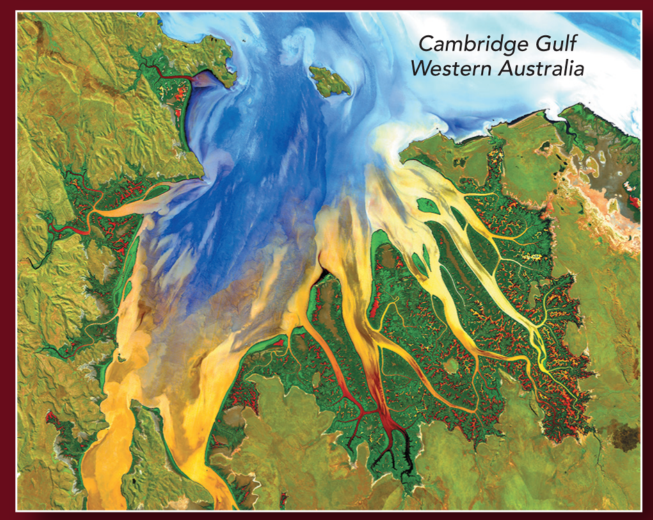
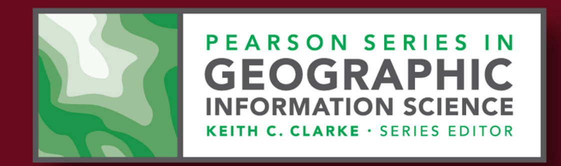
INTRODUCTORY
DIGITAL IMAGE PROCESSING
A Remote Sensing Perspective

4th Edition





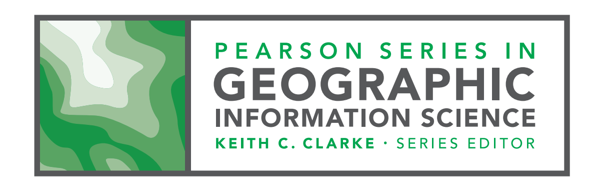




John R. Jensen

INTRODUCTORY
DIGITAL IMAGE PROCESSING
A Remote Sensing Perspective





Berlin/Avery, *Fundamentals of Remote Sensing and Airphoto Interpretation,* 5th edition

Clarke, *Getting Started with Geographic Information Systems,* 5th edition

Clarke, *Maps & Web Mapping*

Greene/Pick, *Exploring the Urban Community: A GIS Approach,* 2nd edition

Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspective,* 4th edition

Jensen, *Remote Sensing of the Environment: An Earth Resource Perspective,* 2nd edition

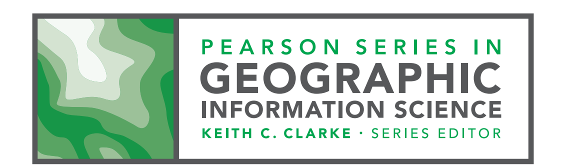
Jensen/Jensen, *Introductory Geographic Information Systems*

Lo/Yeung, *Concepts and Techniques in Geographic Information Systems,* 2nd edition

Slocum/McMaster/Kessler/Howard, *Thematic Cartography and Geovisualization,* 3rd edition

INTRODUCTORY
DIGITAL IMAGE PROCESSING
A Remote Sensing Perspective

John R. Jensen
University of South Carolina



Senior Acquisitions Editor: Christian Botting
Executive Marketing Manager: Neena Bali
Program Manager: Anton Yakovlev
Project Manager: Crissy Dudonis
Editorial Assistant: Amy DeGenaro
Senior Project Manager, Text and Images: Rachel Youdelman
Program Management Team Lead: Kristen Flatham
Project Management Team Lead: David Zielonka
Production Management/Composition: Raghavi Khullar/Cenveo® Publisher Services
Design Manager: Marilyn Perry
Operations Specialist: Maura Zaldivar-Garcia

Cover Photo Credit: **Landsat 8 artist rendition:** NASA, 2014, "*NASA-USGS Landsat 8 Satellite Celebrates First Year of Success,*" February 11, 2014, NASA Goddard Space Flight Center; http://www.nasa.gov/content/goddard/nasa-usgs-landsat-8-satellite-celebrates-first-year-of-success/. **Landsat 8 image of Cambridge Gulf, Northwestern Australia:** U.S. Geological Survey, *Landsat 8 Gallery;* http://landsat.usgs.gov/gallery_view.php?category=nocategory&thesort=pictureId; and, Geoscience Australia; http://www.ga.gov.au/. **Landsat 8 image of lower San Francisco Bay, California, USA:** *NASA Earth Observatory* on May 31, 2013; http://earthobservatory.nasa.gov/IOTD/view.php?id=81238.

Credits and acknowledgments for material borrowed or adapted from other sources and reproduced, with permission, in this textbook appear on the pages at the end of the figure and table captions and in the Acknowledgements.

Jensen, John R., 1949-
Introductory digital image processing : a remote sensing perspective / John R. Jensen, University of South Carolina.
pages cm. -- (Pearson series in geographic information science)
Includes bibliographical references and index.
ISBN 978-0-13-405816-0
1. Remote sensing. 2. Image processing--Digital techniques. I. Title. II. Title: Digital image processing.
G70.4.J46 2015
006.4'2--dc23
2015007022

1 2 3 4 5 6 7 8 9 10—V082—15 14 13 12 11

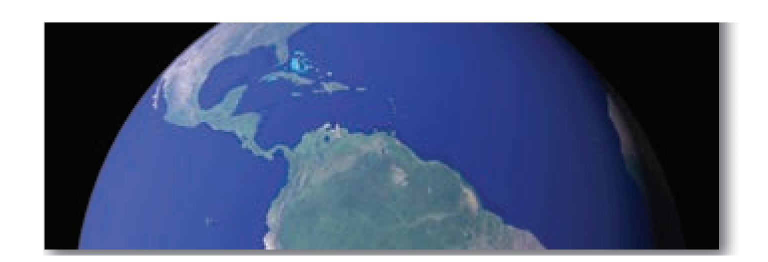


www.pearsonhighered.com

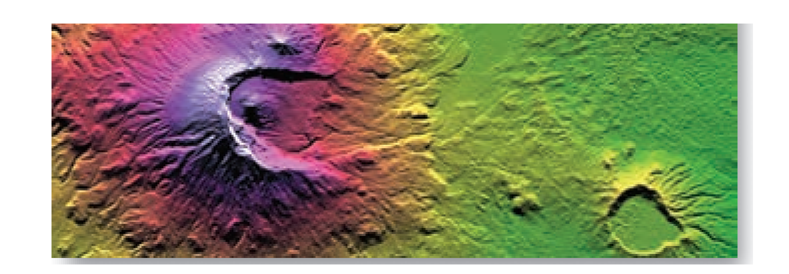
ISBN-10: 0-1340-5816-X
ISBN-13: 978-0-134-05816-0

Brief Contents

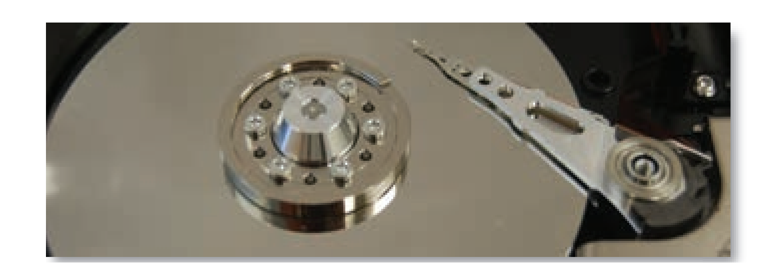
Contents



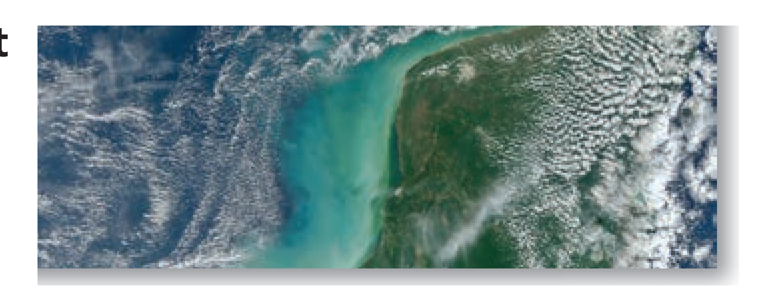
2 Remote Sensing Data Collection



3 Digital Image Processing Hardware and Software



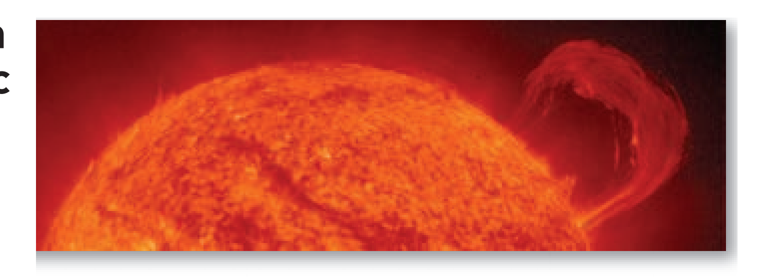
4 Image Quality Assessment and Statistical Evaluation



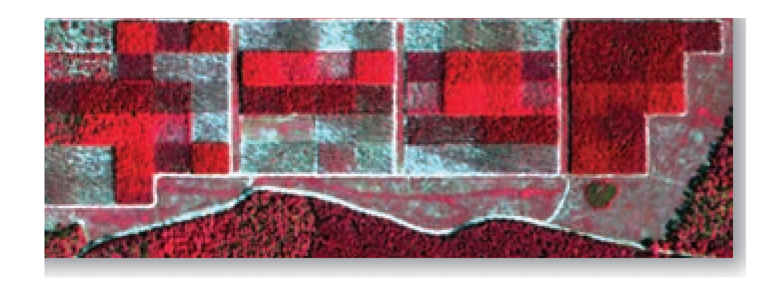
5 Display Alternatives and Scientific Visualization



6 Electromagnetic Radiation Principles and Radiometric Correction



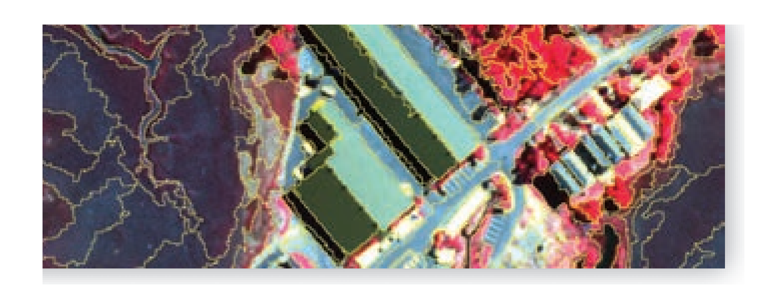
7 Geometric Correction



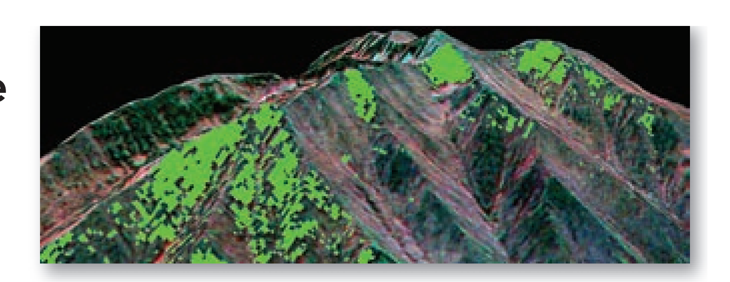
8 Image Enhancement



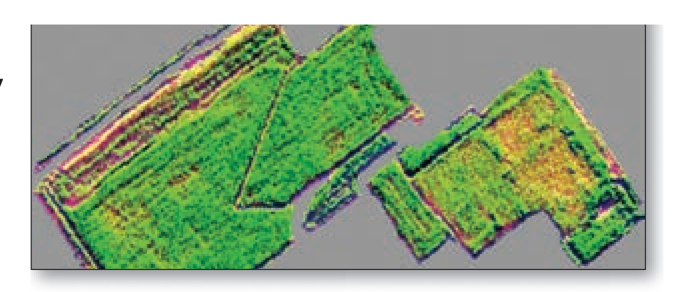

9 Thematic Information Extraction: Pattern Recognition



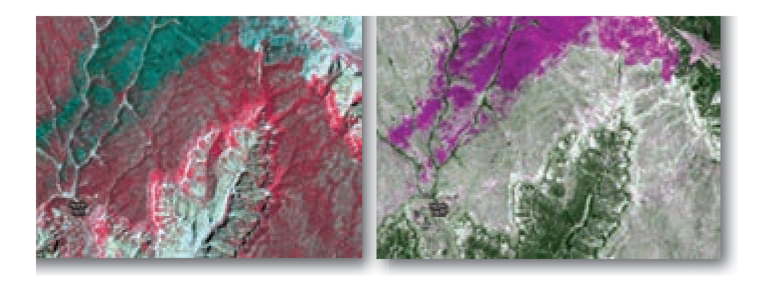
10 Information Extraction Using Artificial Intelligence



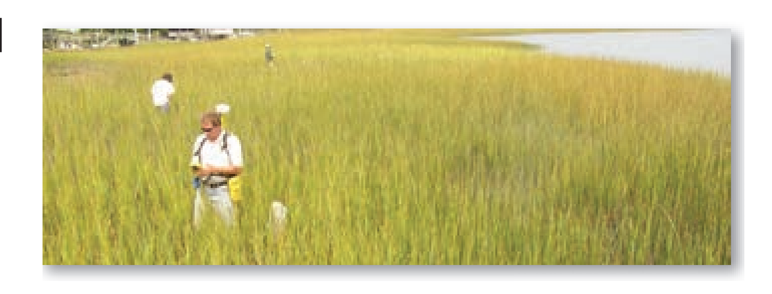
11 Information Extraction Using Imaging Spectroscopy



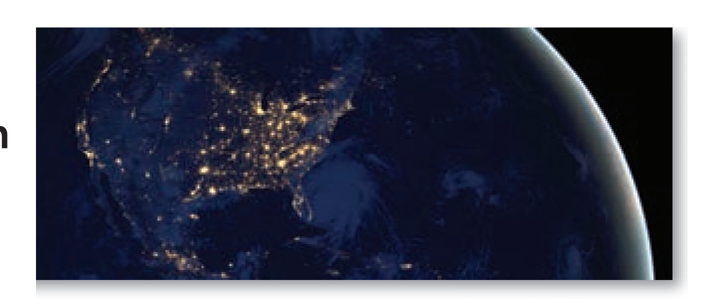
12 Change Detection



13 Remote Sensing-Derived Thematic Map Accuracy Assessment



Appendix: Sources of Imagery and Other Geospatial Information



Preface

New to the 4th Edition

The 4th edition of *Introductory Digital Image Processing: A Remote Sensing Perspective* provides up-to-date information on analytical methods used to analyze digital remote sensing data. The book is ideal for introductory through graduate level university instruction in departments that routinely analyze airborne and/or satellite remote sensor data to solve geospatial problems, including: geography, geology, marine science, forestry, anthropology, biology, soil science, agronomy, urban planning, and others.

Introductory Digital Image Processing (4th edition) is now in full-color. This is important because many digital image processing concepts are best described and demonstrated using a) color diagrams and illustrations, b) the original digital remote sensing displayed in color, and c) digital image processing-derived color images and thematic maps. The reader can now examine the color illustrations and images at their appropriate location in the book rather than having to go to a special signature of color plates.

The goal of this book has always been to take relatively sophisticated digital image processing methods and algorithms and make them as easy to understand as possible for students and for remote sensing scientists. Therefore, the reader will notice that each chapter contains a substantial reference list that can be used by students and scientists as a starting place for their digital image processing project or research. A new appendix provides sources of imagery and other geospatial information. Below is a summary of the major content and improvements in the fourth edition.

Chapter 1: Remote Sensing and Digital Image Processing

Greater emphasis is now placed on the importance of ground reference information that can be used to calibrate remote sensor data and assess the accuracy of remote sensing-derived products such as thematic maps. The "Remote Sensing Process" has been updated to reflect recent innovations in digital image processing. Greater emphasis is now placed on the use of remote sensing to solve local, high-spatial resolution problems as well as for use in global climate change research. This chapter now includes detailed information about the increasing demand for people trained in remote sensing digital image processing. Information is provided from a) the NRC (2013) *Future U.S. Workforce for Geospatial Intelligence* study, and b) U.S. Department of Labor Employment and Training Administration (USDOLETA, 2014) data about the 39,900 "Remote Sensing Scientists and Technologists" and "Remote Sensing Technicians" job openings projected from 2012–2022. Many of these geospatial occupations require training in remote sensing digital image processing.

Chapter 2: Remote Sensing Data Collection

This chapter provides information about historical, current, and projected sources of remotely sensed data. Detailed information about new and proposed satellite remote sensing systems (e.g., Astrium's Pleiades and SPOT 6; DigitalGlobe's GeoEye-1, GeoEye-2, WorldView-1, WorldView-2, WorldView-3; India's CartoSat and ResourceSat; Israel's EROS A2; Korea's KOMPSAT; NASA's Landsat 8; NOAA's NPOESS; RapidEye, etc.) and airborne remote sensing systems (e.g., PICTOMETRY, Microsoft's UltraCAM, Leica's Airborne Digital System 80) are included in the fourth edition. Technical details about decommissioned (e.g., SPOT 1, 2; Landsat 5), degraded (e.g., Landsat 7 ETM^{+}) or failed (e.g., European Space Agency Envisat) sensor systems are provided.

Chapter 3: Digital Image Processing Hardware and Software

As expected, the computer hardware (e.g., CPUs, RAM, mass storage, digitization technology, displays, transfer/storage technology) and software [e.g., multispectral, hyperspectral, per-pixel, object-based image analysis (OBIA)] necessary to perform digital image processing have progressed significantly since the last edition. Improvements in computer hardware often used to perform digital image processing are discussed. The most important functions, characteristics and

sources of the major digital image processing software are provided.

Chapter 4: Image Quality Assessment and Statistical Evaluation

Basic digital image processing mathematical notation is reviewed along with the significance of the histogram. The importance of metadata is introduced. Visual methods of assessing image quality are presented including three-dimensional representation. Univariate and multivariate methods of assessing the initial quality of digital remote sensor data are refreshed. A new section on geostatistical analysis, autocorrelation, and kriging interpolation is provided.

Chapter 5: Display Alternatives and Scientific Visualization

New information is provided on: liquid crystal displays (LCD), image compression alternatives, color coordinate systems (RGB, Intensity-Hue-Saturation, and Chromaticity), the use of 8- and 24-bit color look-up tables, and new methods of merging (fusing) different types of imagery (e.g., Gram-Schmidt, regression Kriging). Additional information is provided about measuring distance, perimeter, shape, and polygon area using digital imagery.

Chapter 6: Electromagnetic Radiation Principles and Radiometric Correction

Additional information is provided about electromagnetic radiation principles (e.g., Fraunhofer absorption features) and the spectral reflectance characteristics of selected natural and human-made materials. Updated information about the most important radiometric correction algorithms is provided, including: a) those that perform absolute radiometric correction (e.g., ACORN, FLAASH, QUAC, ATCOR, empirical line calibration) and, b) those that perform relative radiometric correction (e.g., single and multiple-date image normalization).

Chapter 7: Geometric Correction

Traditional as well as improved methods of image-to-map rectification and image-to-image registration are provided. In addition, this edition contains an expanded discussion on developable surfaces and the properties and advantages/disadvantages of several of the most heavily used cylindrical, azimuthal, and conical map projections. MODIS satellite imagery is projected using selected map projections (e.g., Mercator, Lambert Azimuthal Equal-area). The image mosaicking section contains new examples and demonstrates the characteristics of the USGS annual mosaic of Landsat ETM^+ data (i.e., the *WELD: Web-enabled Landsat Data* project).

Chapter 8: Image Enhancement

The image magnification and reduction sections are revised. In addition, the following image enhancement techniques are updated: band ratioing, neighborhood raster operations, spatial convolution filtering and edge enhancement, frequency filtering, texture extraction, and Principal Components Analysis (PCA). The vegetation indices (VI) section has been significantly revised to include new information on the dominant factors controlling leaf reflectance and the introduction of numerous new indices with graphic examples. Several new texture transforms are introduced (e.g., Moran's *I* Spatial Autocorrelation) and new information is provided on the extraction of texture from images using Grey-level Co-occurrence Matrices (GLCM). The chapter concludes with a new discussion on landscape ecology metrics that can be extracted from remotely sensed data.

Chapter 9: Thematic Information Extraction: Pattern Recognition

Updated information on the American Planning Association Land-Based Classification Standard (NLCS), the U.S. National Land Cover Database (NLCD) Classification System, NOAA's Coastal Change Analysis Program (C-CAP) Classification Scheme, and the IGBP Land-Cover Classification System is included. New methods of feature (band) selection are introduced (e.g., Correlation Matrix Feature Selection). Additional information is provided on Object-Based Image Analysis (OBIA) classification methods, including new OBIA application examples.

Chapter 10: Information Extraction Using Artificial Intelligence

New information is provided on image classification using machine-learning decision trees, regression trees, Random Forest (trees), and Support Vector Machines (SVM). Detailed information is now provided on a number of machine-learning, data-mining decision tree/regression tree programs that can be used to devel-

op production rules (e.g., CART, S-Plus, R Development Core Team, C4.5, C5.0, Cubist). New information about advances in neural network analysis of remote sensor data is included for Multi-layer Perceptrons, Kohonen's Self-Organizing Map, and fuzzy ARTMAP neural networks. A new discussion about the advantages and disadvantages of artificial neural networks is provided.

Chapter 11: Information Extraction Using Imaging Spectroscopy

Advances in airborne and satellite hyperspectral data collection are discussed. Advances in the methods used to process and analyze hyperspectral imagery are provided, including: end-member selection and analysis, mapping algorithms, Spectral Mixture Analysis (SMA), continuum removal, spectroscopic library matching techniques, machine-learning hyperspectral analysis techniques, new hyperspectral indices, and derivative spectroscopy.

Chapter 12: Change Detection

This book has always contained detailed digital change detection information. New information is provided on the impact of sensor system look angle and amount of tree or building obscuration. Advances in binary "change/no-change" algorithms are provided including new analytical methods used to identify the change thresholds and new commercial change detection products such as ESRI's *Change Matters* and MDA's *National Urban Change Indicator*. Significant advances in thematic "from-to" change detection algorithms are discussed including photogrammetric and LiDAR-grammetric change detection, OBIA post-classification comparison change detection, and Neighborhood Correlation Image (NCI) change detection.

Chapter 13: Remote Sensing-Derived Thematic Map Accuracy Assessment

There is a significant amount of literature and debate about the best method(s) to use to determine the accuracy of remote sensing-derived thematic map produced from a single date of imagery or a thematic map derived from multiple dates of imagery (i.e., change detection). The accuracy assessment alternatives and characteristics of the debate are discussed more thoroughly.

Appendix: Sources of Imagery and Other Geospatial Information

Remote sensing data is analyzed best when used in conjunction with other geospatial information. To this end, a new appendix is provided that contains a list of selected geospatial datasets that can be evaluated and/or downloaded via the Internet, including: digital elevation information, hydrology, land use/land cover and biodiversity/habitat, road network and population demographic data, and several types of publicly- and commercially-available remote sensor-data. Map or image examples of the datasets are presented where appropriate.

About Our Sustainability Initiatives

Pearson recognizes the environmental challenges facing this planet, as well as acknowledges our responsibility in making a difference. This book is carefully crafted to minimize environmental impact. The binding, cover, and paper come from facilities that minimize waste, energy consumption, and the use of harmful chemicals. Pearson closes the loop by recycling every out-of-date text returned to our warehouse.

Along with developing and exploring digital solutions to our market's needs, Pearson has a strong commitment to achieving carbon-neutrality. As of 2009, Pearson became the first carbon- and climate-neutral publishing company. Since then, Pearson remains strongly committed to measuring, reducing, and offsetting our carbon footprint.

The future holds great promise for reducing our impact on Earth's environment, and Pearson is proud to be leading the way. We strive to publish the best books with the most up-to-date and accurate content, and to do so in ways that minimize our impact on Earth.

To learn more please visit: **https://www.pearson.com/social-impact.html**.

Acknowledgements

The author appreciates the help provided by the following people:

- Marvin Bauer, Editor-in-Chief, *Remote Sensing of Environment*, Minneapolis, MN.
- Paul Beaty, ERDAS, Intergraph Inc., Atlanta, GA.
- Jigyasa Bhatia, Project Manager, Cenveo Publisher Services, Inc., Noida, India.
- Christian Botting, Senior Editor: Geography, Meteorology, Geospatial Technologies, Pearson Arts & Sciences, Boston, MA.
- Pat Bresnahan, Director, Richland County GIS Division, Columbia, SC.
- Keith Clarke, Department of Geography, University of California at Santa Barbara, CA, reviewer.
- Leslie Godwin, Geography Division, U.S. Bureau of the Census.
- Russ Congalton, Department of Natural Resources & the Environment, University of New Hampshire, reviewer.
- John Copple, CEO, Sanborn Map Company, Inc., Colorado Springs, CO.
- Dave Cowen, Department of Geography, University of South Carolina, Columbia, SC.
- John Dykstra, MDA Information Systems, LLC.
- Tony Filippi, Department of Geography, Texas A&M University, reviewer.
- Brian Hayman, Publisher, *GIScience & Remote Sensing*, Bellwether Publishing, Columbia, MD.
- Mike Hodgson, Department of Geography, University of South Carolina, Columbia, SC.
- Jungho Im, Ulsan National Institute of Science and Technology, Korea, reviewer.
- Ryan Jensen, Dept. of Geography, Brigham Young University, Provo, UT.
- Raghavi Khullar, Associate Project Manager, Cenveo Publisher Services, Inc., Noida, India
- Charlie Mondello, Executive Vice-president, PICTOMETRY International, Inc., Rochester, NY.
- Dan Morgan, Director, GIS Department, Beaufort County, SC.
- George Raber, Department of Geography, University of Southern Mississippi, reviewer.
- Kevin Remington, Department of Geography, University of South Carolina, Columbia, SC.
- Lynn Shirley, Department of Geography, University of South Carolina, Columbia, SC.
- Doug Stow, Department of Geography, California State University at San Diego, reviewer.
- Jason Tullis, Department of Geography, University of Arkansas, reviewer.

The following individuals, agencies, and commercial firms provided data, imagery, and/or software used to create the illustrations:

- Airbus Defense & Space®, Inc.: information about SPOT, Pleiades, and other satellite remote sensing systems.
- American Society for Photogrammetry & Remote Sensing, Bethesda, MD: selected figures published in *Photogrammetric Engineering & Remote Sensing*.
- Beaufort County, SC, GIS Department: digital aerial photography and parcel data.
- Bellwether Publishing, Ltd.: selected figures published in *GIScience & Remote Sensing*.
- Blackbridge®, Inc.: information about RapidEye satellite products.
- DigitalGlobe®, Inc.: use of IKONOS, GeoEye-1, QuickBird, Worldview-1, Worldview-2, and Worldview-3 satellite imagery.
- DMC International Imaging®, Ltd.: information about their imaging systems.
- Elsevier Publishing, Ltd.: selected figures published in *Remote Sensing of Environment*.
- Esri, Inc.: ArcGIS® ArcMap® and ArcCatalog® graphical user interfaces are the intellectual property of Esri and are used herein by permission; information about ChangeMatters® software. All rights reserved.
- Exelis®, Inc.: information about ENVI®, QUAC®, and FLAASH® image processing software and graphical user interfaces.
- Geoscience Australia: enhanced Landsat 8 image of northwestern Australia.
- Geosystems®, GmbH: images and information about ATCOR® atmospheric correction software.

- Google®, Inc.: photograph of *Google earth*® interface.
- Hexagon Geospatial®, Inc., and Intergraph®, Inc.: information about ERDAS Imagine® image processing software and graphical user interface.
- HyVista®, Inc.: information about the HyMap® hyperspectral sensor system.
- ImageSat International®, Inc.: information about EROS A and EROS B.
- Indian Space Research Organization (ISRO) and the National Remote Sensing Centre (NRSC): information about various imaging systems.
- Intel®, Inc.: data and information about microprocessors.
- Itres Research®, Ltd., of Canada: information about their remote sensing systems.
- Korean Aerospace Research Institute (KARI): information about various imaging systems.
- Leica Geosystems®, Inc.: information about imaging and LiDAR sensor systems and the Leica Photogrammetry Suite® software.
- MDA Information Systems®, LLC.: information about the National Urban Change Indicator (NUCI)® software.
- Microsoft®, Inc.: information about the UltraCam Eagle® camera and *bing*® search engine.
- National Aeronautics & Space Administration (NASA) and Jet Propulsion Laboratory (JPL): ASTER, AVIRIS, G-Projector software, Landsat Thematic Mapper and ETM^+, Landsat 8, MODIS, SRTM.
- National Archives and Records Administration (NARA): gray-scale and color cards.
- National Oceanic and Atmospheric Administration (NOAA): AVHRR, Coastal Change Analysis Program (C-CAP), NOAA Geophysical Data System (GEODAS).
- Oak Ridge National Laboratory (ORNL): LandScan description.
- PICTOMETRY International®, Inc. and Eagleview®, Inc.; digital vertical and oblique aerial photography and ChangeFindr® software.
- Sanborn Map Company®, Inc.: digital aerial photography.
- Sensefly®, LLC.: unmanned aerial vehicle (UAV) photograph.
- State of Utah, Automated Geographic Reference Center (AGRC): aerial photography of Utah.
- Taylor & Francis Publishing, Ltd.: selected figures published in *International Journal of Remote Sensing*, *Geocarto International*, and *GIScience & Remote Sensing*.
- U.S. Bureau of the Census: for permission to use Jensen et al. (2012) report on change detection.
- U.S. Department of Agriculture (USDA): National Agricultural Imagery Program (NAIP).
- U.S. Fish & Wildlife Service (USF&WS): National Wetlands Inventory (NWI) classification scheme.
- U. S. Geological Survey (USGS): Declassified satellite imagery, Digital Orthophoto Quadrangles (DOQ), *Earth Explorer* website, Elevation Derivatives for National Applications (EDNA), GAP Analysis Program, *Data.gov* website, Global Visualization Viewer (GLOVIS) website, GTOPO30, historical terrestrial photographs, Landsat 8, LiDAR, National Map, National Elevation Dataset (NED), National Hydrography Dataset (NHD), National Land Cover Dataset (NLCD), Map Projections, US Topo data, Web-enabled Landsat Data (WELD).
- The Nature Conservancy: permission to use sponsored research results.

I appreciate the thoughtful editorial support provided by Christian Botting (Senior Editor: Geography, Meteorology, Geospatial Technologies) and Crissy Dudonis of Pearson, Inc. Raghavi Khullar and Jigyasa Bhatia of Cenveo Publishing Services provided quality production support.

Special thanks goes to my wife, Marsha, for her continuous encouragement, sacrifice, and help.

John R. Jensen
University of South Carolina

About the Author

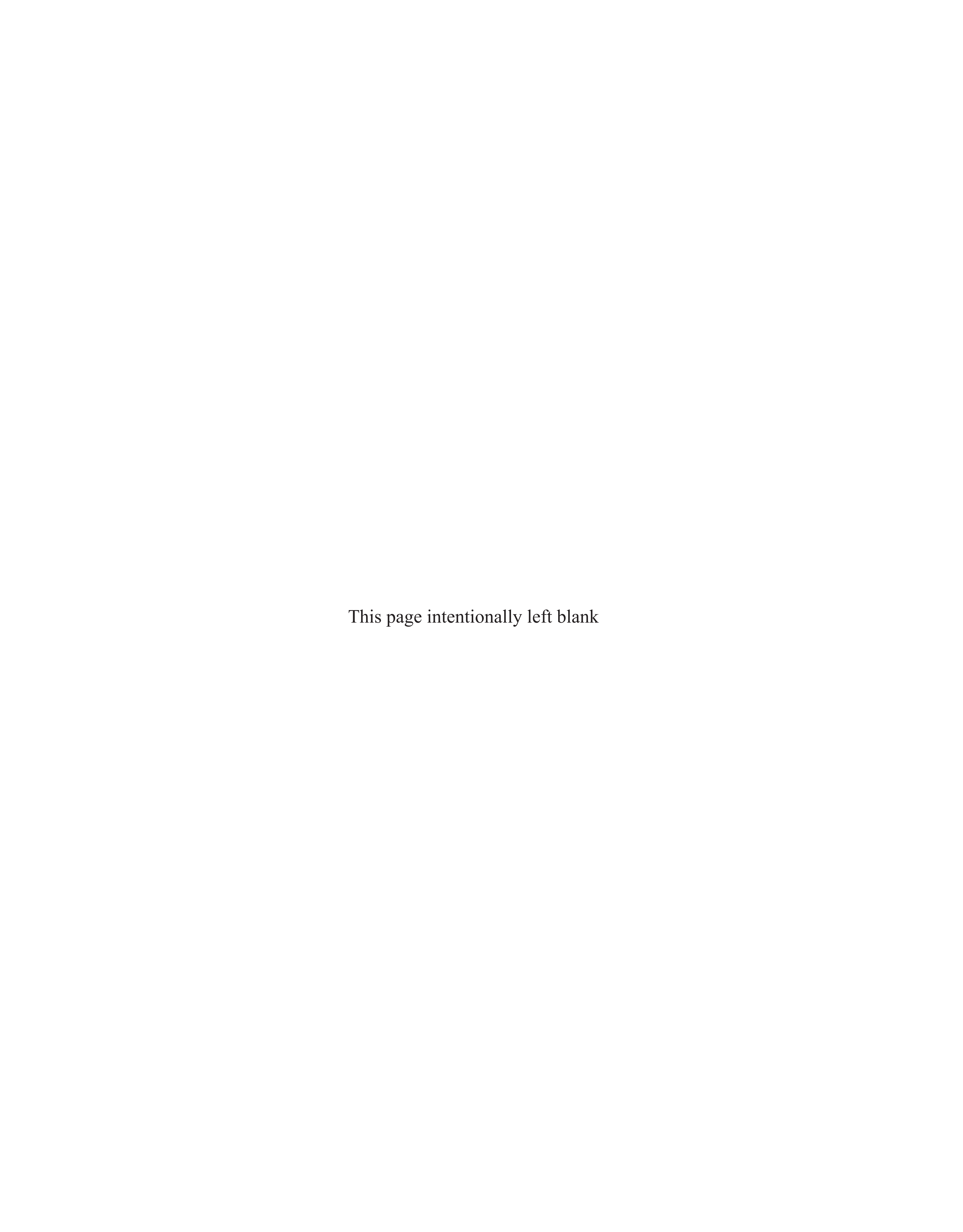
John R. Jensen received a BA in geography from California State University at Fullerton, an MS from Brigham Young University (BYU), and a PhD from the University of California at Los Angeles (UCLA). He is a Carolina Distinguished Professor Emeritus in the Department of Geography at the University of South Carolina. He is a certified photogrammetrist and a past president of the American Society for Photogrammetry & Remote Sensing (ASP&RS): The Geospatial Information Society.

Dr. Jensen has conducted more than 50 remote sensing-related research projects sponsored by NASA, DOE, NOAA, U.S. Bureau of the Census, and the Nature Conservancy and published more than 120 refereed journal articles. He has mentored 34 Ph.D. and 62 master's students.

Dr. Jensen received the *SAIC/ASP&RS John E. Estes Memorial Teaching Award* for education, mentoring, and training in remote sensing and GIS. He received the U.S. Geological Survey (USGS)/National Aeronautics & Space Administration (NASA) *William T. Pecora Award* for his remote sensing research contributions. He received the Association of American Geographers (AAG) *Lifetime Achievement Award* for research and education in remote sensing and GIScience. He became an *Honorary Member* of the ASP&RS in 2013.

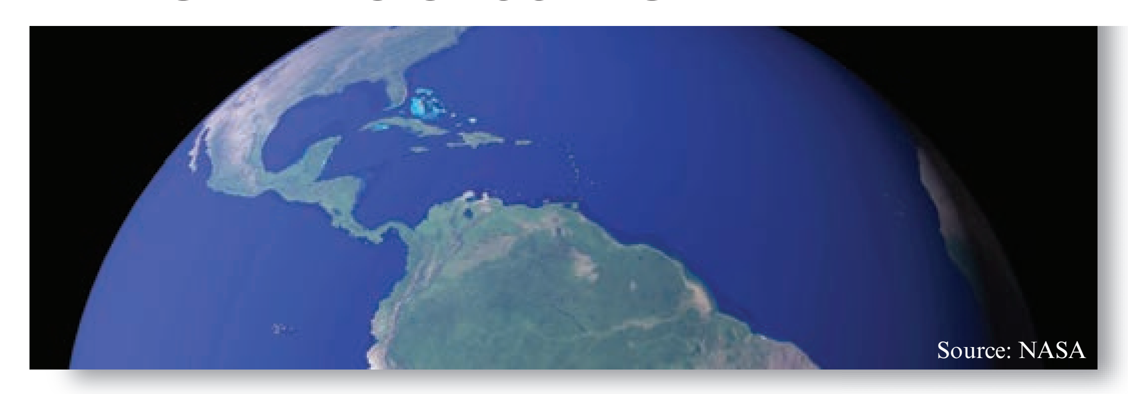
He has served on numerous editorial boards and was the Editor-in-chief of the journal *GIScience & Remote Sensing* now published by Taylor & Francis. He is co-author of *Introductory Geographic Information Systems* and author of *Remote Sensing of the Environment: An Earth Resource Perspective*, 2nd edition, also published by Pearson.

Dr. Jensen has been associated with eight National Research Council (NRC) remote sensing-related committees and subsequent National Academy Press publications.

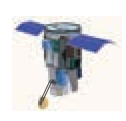


This page intentionally left blank

1 REMOTE SENSING AND DIGITAL IMAGE PROCESSING

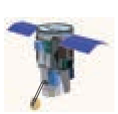

Source: NASA

Scientists observe nature and man-made phenomena, make measurements, and then attempt to test hypotheses concerning these phenomena. The data collection may take place directly in the field (referred to as *in situ* or in-place data collection), and/or at some remote distance from the subject matter (commonly referred to as **remote sensing** of the environment). Remote sensing technology is now used routinely to obtain accurate, timely information for a significant variety of applications, including: the study of daily weather and long-term climate change; urban-suburban land-use/land cover monitoring; ecosystem modeling of vegetation, water, snow/ice; food security; military reconnaissance; and many others (NRC, 2007ab, 2009, 2013, 2014). The majority of the remotely sensed data are analyzed using digital image processing techniques.



Overview

This chapter introduces basic *in situ* data-collection considerations. Remote sensing is then formally defined along with its advantages and limitations. The remote sensing–process is introduced with particular attention given to: a) the statement of the problem, b) identification of *in situ* and remote sensing data requirements, c) remote sensing data collection using satellite and airborne sensor systems, d) the conversion of remote sensing data into information using analog and/or digital image processing techniques, and e) accuracy assessment and information presentation alternatives. Earth observation economics are considered along with remote sensing and digital image processing careers in the public and private sectors. The organization of the book is reviewed along with its earth resource analysis perspective.



In Situ Data Collection

One form of *in situ* data collection involves the scientist questioning the phenomena of interest. For example, a census enumerator may go door to door, asking people questions about their age, sex, education, income, etc. These data are recorded and used to document the demographic characteristics of the population.

Conversely, a scientist may use a transducer or other *in situ* instrument to make measurements. Transducers are usually placed in direct physical contact with the object of interest. Many different types of transducers are available. For example, a scientist could use a thermometer to measure the temperature of the air, soil, or water; an anemometer to measure wind speed; or a psychrometer to measure air humidity. The data recorded by the transducers may be an analog electrical signal with voltage variations related to the intensity of the property being measured. Often these analog signals are transformed into digital values using analog-to-digital (A-to-D) conversion procedures. *In situ* data collection using transducers relieves the scientist of monotonous data collection often in inclement weather. Also, the scientist can distribute the transducers at geographic locations throughout the study area, allowing the same type of measurement to be obtained at many locations at the same time. Sometimes data from the transducers are telemetered electronically to a central collection point for rapid evaluation and archiving.

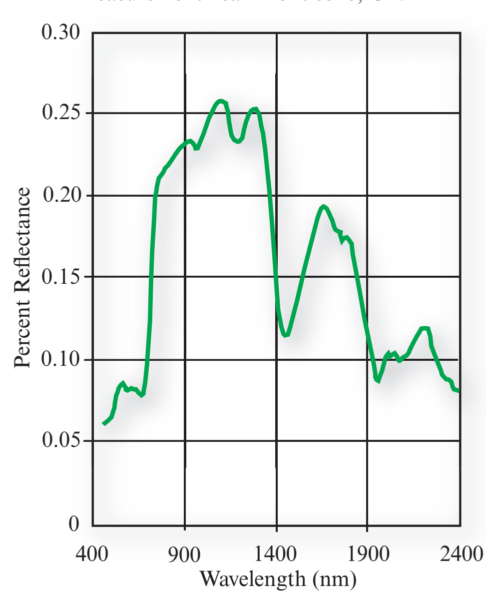
***In Situ* Measurement**



a. Vegetation height and GPS location measurement near Monticello, UT.



b. Spectral reflectance measurement of grassland near Monticello, UT.



c. Spectral reflectance characteristics of the grassland in Figure 1-1b.



d. Above-canopy calibration of a ceptometer used to measure leaf-area-index (LAI) near Monticello, UT.

FIGURE 1–1 *In situ* (in-place) data are collected in the field. a) Vegetation height is being measured using a simple metal ruler and geographic location is measured using a hand-held GPS unit. b) Vegetation spectral reflectance information is being collected using a spectroradiometer held approximately 1 m above the canopy. The *in situ* spectral reflectance measurements may be used to calibrate the spectral reflectance measurements obtained from a remote sensing system. c) Spectral reflectance characteristics of the area on the ground in Figure 1-1b. d) The leaf-area-index (LAI) of the grassland can be measured using a ceptometer that records the number of "sunflecks" that pass through the vegetation canopy. Ceptometer measurements are made just above the canopy as shown and on the ground below the canopy. These measurements are then used to compute *in situ* LAI that can be used to calibrate remote sensing–derived LAI estimates.

Several examples of *in situ* data collection on a grassland near Monticello, UT, are demonstrated in Figure 1-1. A stake is used to mark the location of *in situ* sample "UPL 5" in Figure 1-1a. A hand-held Global Positioning System (GPS) unit is placed at the base of the stake to obtain precise *x,y,* and *z* location information about the sample with a horizontal accuracy of approximately ±0.25 m. A dimensionally-stable metal ruler is used to measure the height of the grass at the sample location. A hand-held spectroradiometer is used to measure the spectral reflectance characteristics of the materials within the Instantaneous-Field-Of-View (IFOV) of the radiometer on the ground (Figure 1-1b). The spectral reflectance characteristics of the grassland within the IFOV from 400 – 2,400 nm are shown in Figure 1-1c. A hand-held ceptometer is being used to collect information about incident skylight above the vegetation canopy in Figure 1-1d. The scientist will then place the ceptometer on the ground beneath the vegetation canopy present. Approximately 80 photo-diodes along the linear ceptometer measure the amount of light that makes its way through the vegetation canopy to the ceptometer. The above- and below-canopy ceptometer measurements are used to compute the Leaf-Area-Index (LAI) at the sample location. The greater the amount of vegetation, the greater the LAI. Interestingly, all of these measurement techniques are non-destructive (i.e., vegetation clipping or harvesting is not required).

Data collection by scientists in the field or by instruments placed in the field provides much of the data for physical, biological, and social science research. However, it is important to remember that no matter how careful the scientist is, error may be introduced during the *in situ* data-collection process. First, the scientist in the field can be *intrusive*. This means that unless great care is exercised, the scientist can actually change the characteristics of the phenomenon being measured during the data-collection process. For example, a scientist collecting a vegetation spectral reflectance reading could inadvertently step on the sample site, disturbing the vegetation prior to data collection.

Scientists may also collect data in the field using biased procedures. This introduces *method-produced error*. It could involve the use of a biased sampling design or the systematic, improper use of a piece of equipment. Finally, the *in situ* data-collection measurement device may be calibrated incorrectly. This can result in serious *measurement-device calibration error*.

Intrusive *in situ* data collection, coupled with human method-produced error and measurement-device miscalibration, all contribute to *in situ* data-collection error. Therefore, it is a misnomer to refer to *in situ* data as ground truth data. Instead, we should simply refer to it as *in situ* **ground reference data**, acknowledging that it contains error.



Remote Sensing Data Collection

Fortunately, it is possible to collect information about an object or geographic area from a distant vantage point using **remote sensing** instruments (Figure 1-2). Remote sensing data collection was originally performed using cameras mounted in suborbital aircraft. **Photogrammetry** was defined in the early editions of the *Manual of Photogrammetry* as:

> the art or science of obtaining reliable measurement by means of photography (American Society of Photogrammetry, 1952; 1966).

Photographic interpretation is defined as:

> the act of examining photographic images for the purpose of identifying objects and judging their significance (Colwell, 1960).

Remote sensing was formally defined by the American Society for Photogrammetry and Remote Sensing (ASPRS) as:

> the measurement or acquisition of information of some property of an object or phenomenon, by a recording device that is not in physical or intimate contact with the object or phenomenon under study (Colwell, 1983).

In 1988, ASPRS adopted a **combined definition of photogrammetry and remote sensing**:

> Photogrammetry and remote sensing are the art, science, and technology of obtaining reliable information about physical objects and the environment, through the process of recording, measuring and interpreting imagery and digital representations of energy patterns derived from non-contact sensor systems (Colwell, 1997).

But where did the term *remote sensing* come from? The actual coining of the term goes back to an unpublished paper in the early 1960s by the staff of the Office of Naval Research (ONR) Geography Branch (Pruitt, 1979; Fussell et al., 1986). Evelyn L. Pruitt was the author of the paper. She was assisted by staff member Walter H. Bailey. Aerial photo interpretation had become very important in World War II. The space age was just getting under way with the 1957 launch of *Sputnik* (U.S.S.R.), the 1958 launch of *Explorer 1*

(U.S.), and the collection of photography from the then secret CORONA program initiated in 1960 (Table 1-1, page 11). In addition, the Geography Branch of ONR was expanding its research using instruments other than cameras (e.g., scanners, radiometers) and into regions of the electromagnetic spectrum beyond the visible and near-infrared regions (e.g., thermal infrared, microwave). Thus, in the late 1950s it had become apparent that the prefix "photo" was being stretched too far in view of the fact that the root word, *photography*, literally means "to write with [visible] light" (Colwell, 1997). Evelyn Pruitt (1979) wrote:

> The whole field was in flux and it was difficult for the Geography Program to know which way to move. It was finally decided in 1960 to take the problem to the Advisory Committee. Walter H. Bailey and I pondered a long time on how to present the situation and on what to call the broader field that we felt should be encompassed in a program to replace the aerial photointerpretation project. The term 'photograph' was too limited because it did not cover the regions in the electromagnetic spectrum beyond the 'visible' range, and it was in these nonvisible frequencies that the future of interpretation seemed to lie. 'Aerial' was also too limited in view of the potential for seeing the Earth from space.

The term **remote sensing** was promoted in a series of symposia sponsored by ONR at the Willow Run Laboratories of the University of Michigan in conjunction with the National Research Council throughout the 1960s and early 1970s, and has been in use ever since (Estes and Jensen, 1998).

Remote sensing instruments such as cameras, multispectral and hyperspectral sensors, thermal-infrared detectors, Radio Detection and Ranging (RADAR) sensors, and Light Detection and Ranging (LiDAR) instruments are flown onboard satellites or suborbital aircraft such as airplanes, helicopters, and Unmanned Aerial Vehicles (UAVs) (Figure 1-2). Sound Navigation and Ranging (SONAR) sensors are placed onboard ships and submarines to map the bathymetry of subsurface terrain.

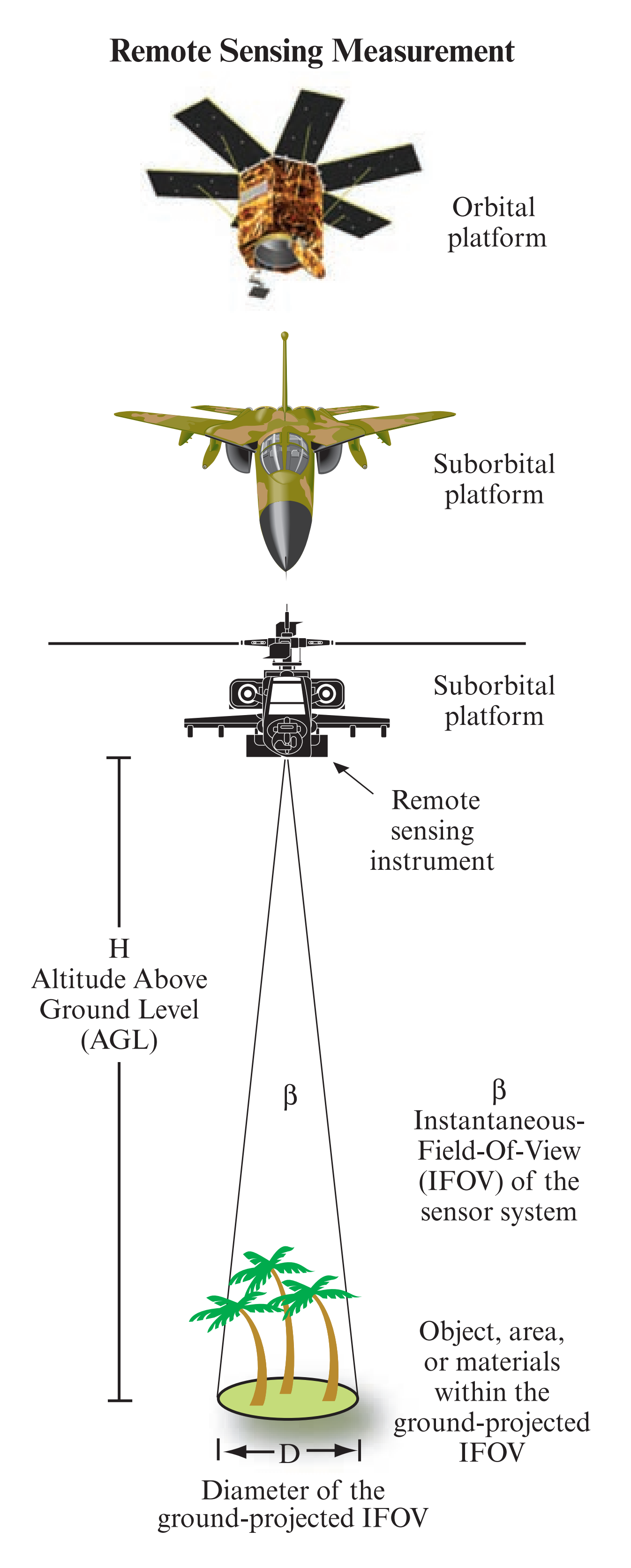


FIGURE 1–2 A remote sensing instrument collects information about an object or phenomenon within the IFOV of the sensor system without being in direct physical contact with it. The remote sensing instrument may be located just a few meters above the ground or onboard an aircraft or satellite platform.

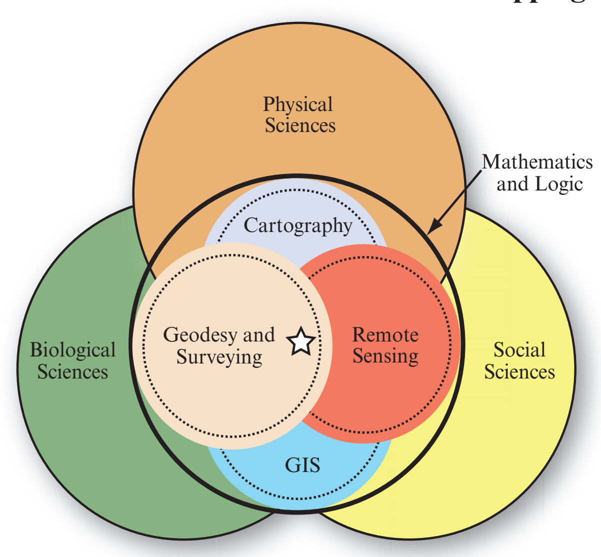
Observations About Remote Sensing

The following brief discussion focuses on various terms found in the formal definitions of remote sensing.

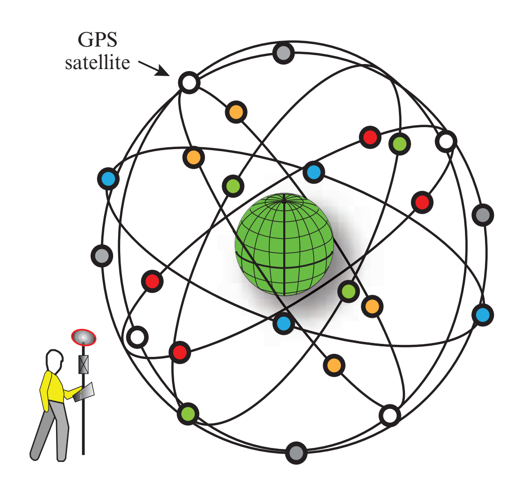
Remote Sensing: Art and/or Science?

Science: A **science** is a broad field of human knowledge concerned with facts held together by principles (rules). Scientists discover and test facts and principles

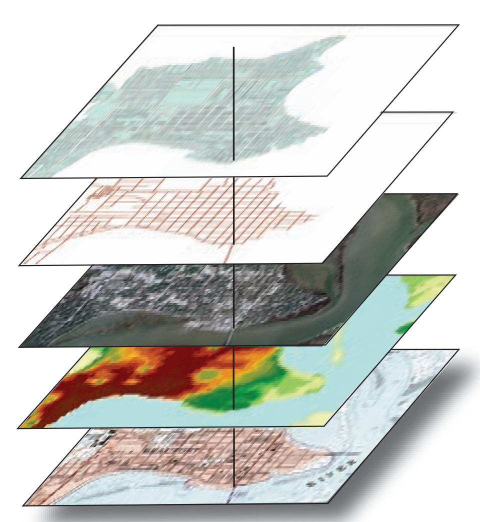
Allied Mapping Sciences



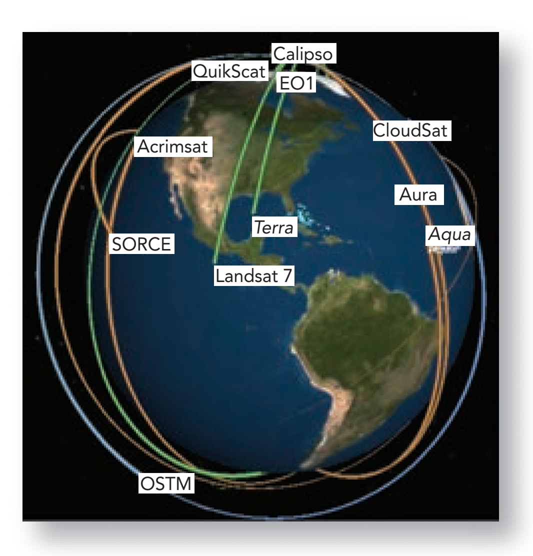
a. Interaction of the geographic information sciences (GISciences) as they relate to the other sciences.



b. Geodesy and surveying.



c. Cartography and Geographic Information Systems (GIS).



d. Remote sensing.

FIGURE 1–3 a) Interaction model depicting the relationship of the geographic information sciences (geodesy, surveying, cartography, GIS, and remote sensing) as they relate to mathematics and logic and the physical, biological, and social sciences. b) Geodesy and surveying provide geodetic control and detailed local geospatial information. c) Cartographers map geospatial information and Geographic Information Systems are used to model geospatial information. d) Satellite and suborbital remote sensing systems provide much of the useful geospatial information used by cartographers, GIS practitioners, and other scientists.

by the scientific method, an orderly system of solving problems. Scientists generally feel that any subject that humans can study by using the scientific method and other special rules of thinking may be called a science. The sciences include 1) *mathematics and logic*, 2) *physical sciences*, such as physics and chemistry, 3) *biological sciences*, such as botany and zoology, and 4) *social sciences*, such as geography, sociology, and anthropology (Figure 1-3a). Interestingly, some persons do not consider mathematics and logic to be sciences. But the fields of knowledge associated with mathematics and logic are such valuable *tools* for science that we cannot ignore them. The human race's earliest questions were concerned with "how many" and "what belonged together." They struggled to count, to classify, to think systematically, and to describe exactly. In many respects, the state of development of a science is indicated by the use it makes of mathematics. A science seems to begin with simple mathematics to measure, then works toward more complex mathematics to explain.

Remote sensing is a tool or technique similar to mathematics. Using sophisticated sensors to measure the amount of electromagnetic energy exiting an object or geographic area from a distance and then extracting valuable information from the data using mathematically and statistically based algorithms is a scientific activity. Remote sensing functions in harmony with several other geographic information sciences (often referred to as the **GISciences**), including geodesy, surveying, cartography, and Geographic Information Systems (GIS) (Figure 1-3b-d). The model shown in Figure 1-3a suggests there is interaction between the mapping sciences, where no subdiscipline dominates and all are recognized as having unique yet overlapping areas of knowledge and intellectual activity as they are used in physical, biological, and social science research.

Significant advances will continue to be made in all of these technologies. In addition, the technologies will become more integrated with one another. For example, GIS network analysis applications have benefited tremendously from more accurate road network centerline data obtained using GPS units mounted on specially prepared cars or derived from high spatial resolution remote sensor data. Entire industries are now dependent on high spatial resolution satellite and airborne remote sensor data as geographic background imagery for their search engines (e.g., *Google Earth*, *Google Maps*, *Bing Maps*). Remote sensing data collection has benefited from advancements in GPS measurements that are used to improve the geometric accuracy of images and image-derived products that are used so heavily in GIS applications. Terrestrial surveying has been revolutionized by improvements in GPS technology, where measurements as accurate as ± 1 to 3 cm in x, y, and z are now possible.

The theory of science suggests that scientific disciplines go through four classic developmental stages. Wolter (1975) suggested that the growth of a scientific discipline, such as remote sensing, that has its own techniques, methodologies, and intellectual orientation seems to follow the sigmoid or logistic curve illustrated in Figure 1-4. The growth stages of a scientific field are: Stage 1—a preliminary growth period with small increments of literature; Stage 2—a period of exponential growth when the number of publications doubles at regular intervals; Stage 3—a period when the rate of growth begins to decline but annual increments remain constant; and Stage 4—a final period when the rate of growth approaches zero. The characteristics of a scholarly field during each of the stages may be briefly described as follows: Stage 1—little or no social organization; Stage 2—groups of collaborators and existence of invisible colleges, often in the form of ad hoc institutes, research units, etc.; Stage 3—increasing specialization and increasing controversy; and Stage 4—decline in membership in both collaborators and invisible colleges.

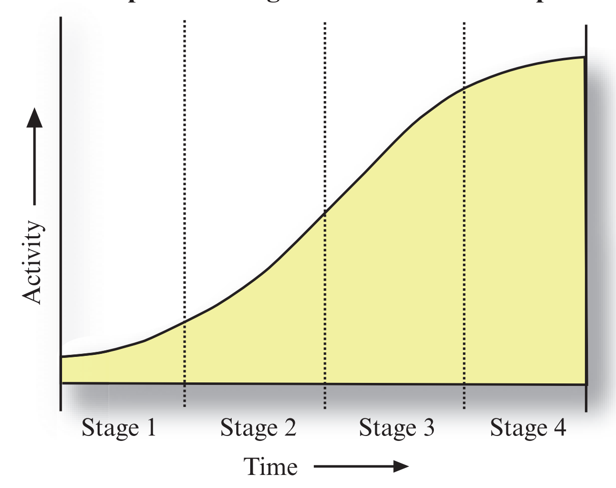


FIGURE 1–4 Developmental stages of a scientific discipline (Source: Wolter, 1975; Jensen and Dahlberg, 1983).

Using this logic, it appears that remote sensing is in Stage 2, experiencing exponential growth since the mid-1960s with the number of publications doubling at regular intervals. Empirical evidence consists of: 1) the organization of many specialized institutes and centers of excellence associated with remote sensing, 2) the organization of numerous professional societies devoted to remote sensing research, 3) the publication of numerous new scholarly remote sensing journals, 4) significant technological advancement such as improved sensor systems and methods of image analysis, and 5) robust self-examination. We may be approaching Stage 3 with increasing specialization and theoretical controversy. However, the rate of growth of remote sensing has not begun to decline. In fact, there has been a tremendous surge in the numbers of persons specializing in remote sensing and commercial firms using remote sensing during the last three decades. Significant improvements in the spatial resolution of satellite remote sensing (e.g., high resolution 1 × 1 m panchromatic data) has brought even more social science GIS practitioners into the fold. Hundreds of new peer-reviewed remote sensing research articles are published every month.

Art: The process of visual photo or image interpretation brings to bear not only scientific knowledge, but all of the background that a person has obtained through his or her lifetime. Such learning cannot be measured, programmed, or completely understood. The synergism of combining scientific knowledge with the real-world analyst experience allows the interpreter to develop heuristic rules of thumb to extract valuable

information from the imagery. It is a fact that some image analysts are superior to other image analysts because they: 1) understand the scientific principles better, 2) are more widely traveled and have seen many landscape objects and geographic areas, and/or 3) can synthesize scientific principles and real-world knowledge to reach logical and correct conclusions. Thus, remote sensing image interpretation is both an art and a science.

Information About an Object or Area

Sensors can obtain very specific information about an object (e.g., the diameter of an oak tree crown) or the geographic extent of a phenomenon (e.g., the polygonal boundary of an entire oak forest). The electromagnetic energy emitted or reflected from an object or geographic area is used as a surrogate for the actual property under investigation. The electromagnetic energy measurements are typically turned into information using visual and/or digital image processing techniques.

The Instrument (Sensor)

Remote sensing is performed using an instrument, often referred to as a sensor. The majority of remote sensing instruments record electromagnetic radiation (EMR) that travels at a velocity of 3×10^8 m s^{-1} from the source, directly through the vacuum of space or indirectly by reflection or reradiation to the sensor. The EMR represents a very efficient high-speed communications link between the sensor and the remote phenomenon. In fact, we know of nothing that travels faster than the speed of light. Changes in the amount and properties of the EMR become, upon detection by the sensor, a valuable source of data for interpreting important properties of the phenomenon (e.g., temperature, color). Other types of force fields may be used in place of EMR, such as acoustic (sonar) waves. However, the majority of remotely sensed data collected for Earth resource applications is the result of sensors that record electromagnetic energy.

Distance: How Far Is Remote?

Remote sensing occurs at a distance from the object or area of interest. Interestingly, there is no clear distinction about how great this distance should be. The intervening distance could be 1 cm, 1 m, 100 m, or more than 1 million m from the object or area of interest. Much of astronomy is based on remote sensing. In fact, many of the most innovative remote sensing systems and visual and digital image processing methods were originally developed for remote sensing extraterrestrial landscapes such as the moon, Mars, Saturn, Jupiter, etc. (especially by NASA's Jet Propulsion Laboratory personnel). This text, however, is concerned primarily with remote sensing of the terrestrial Earth, using sensors that are placed on suborbital air-breathing aircraft or orbital satellite platforms placed in the vacuum of space.

Remote sensing and digital image processing techniques can also be used to analyze inner space. For example, an electron microscope can be used to obtain photographs of extremely small objects on the skin, in the eye, etc. An x-ray instrument is a remote sensing system where the skin and muscle are like the atmosphere that must be penetrated, and the interior bone or other matter is the object of interest.

Remote Sensing Advantages and Limitations

Remote sensing has several unique advantages as well as some limitations.

Advantages

Remote sensing is unobtrusive if the sensor is passively recording the electromagnetic energy reflected from or emitted by the phenomenon of interest. This is a very important consideration, as passive remote sensing does not disturb the object or area of interest.

Remote sensing devices are programmed to collect data systematically, such as within a single 9 × 9 in. frame of vertical aerial photography or a matrix (raster) of Landsat 5 Thematic Mapper data. This systematic data collection can remove the sampling bias introduced in some *in situ* investigations.

Remote sensing science is also different from cartography or GIS because these sciences rely on data obtained or synthesized by others. Remote sensing science can provide fundamental, *new* scientific data or information. Under controlled conditions, remote sensing can provide fundamental biophysical information, including: x, y location, z elevation or depth, biomass, temperature, moisture content, etc. In this sense, remote sensing science is much like surveying, providing fundamental information that other sciences can use when conducting scientific investigations. However, unlike much of surveying, the remotely sensed data can be obtained systematically over very large geographic areas rather than just single-point observations. In fact, remote sensing–derived information is now critical to the successful modeling of numerous natural (e.g., water-supply estimation; eutrophication studies; nonpoint source pollution) and cultural (e.g., land-use conversion at the urban fringe; water-demand estimation; population estimation; food security) processes (NRC, 2007a; 2009). A good example is the digital elevation model that is so important in many spatially-distributed GIS models. Digital elevation models are now produced mainly from light detection and ranging

(LiDAR) (e.g., Renslow, 2012), stereoscopic aerial photography, RADAR measurements, or Interferometric Synthetic Aperture Radar (IFSAR) imagery.

Limitations

Remote sensing science has limitations. Perhaps the greatest limitation is that it is often oversold. Remote sensing is not a panacea that will provide all the information needed to conduct physical, biological, or social science research. It simply provides some spatial, spectral, and temporal information of value in a manner that is hopefully efficient and economical.

Human beings select the most appropriate remote sensing system to collect the data, specify the various resolution(s) of the remote sensor data, calibrate the sensor, select the satellite or suborbital platform that will carry the sensor, determine when the data will be collected, and specify how the data are processed. Therefore, human method-produced error may be introduced as the remote sensing instrument and mission parameters are specified.

Powerful active remote sensor systems that emit their own electromagnetic radiation (e.g., LiDAR, RADAR, SONAR) can be intrusive and affect the phenomenon being investigated. Additional research is required to determine how intrusive these active sensors can be.

Remote sensing instruments may become uncalibrated, resulting in uncalibrated remote sensor data. Finally, remote sensor data may be relatively expensive to collect and analyze. Hopefully, the information extracted from the remote sensor data justifies the expense. The greatest expense in a typical remote sensing investigation is for well-trained image analysts, not remote sensor data.



The Remote Sensing Process

Scientists have been developing procedures for collecting and analyzing remotely sensed data for more than 150 years. The first photograph from an aerial platform (a tethered balloon) was obtained in 1858 by the Frenchman Gaspard Felix Tournachon (who called himself Nadar). Significant strides in aerial photography and other remote sensing data collection took place during World Wars I and II, the Korean Conflict, the Cuban Missile Crisis, the Vietnam War, the Gulf War, the war in Bosnia, and the war on terrorism. Basically, military contracts to commercial companies resulted in the development of sophisticated electro-optical multispectral remote sensing systems and thermal infrared and microwave (radar) sensor systems whose characteristics are summarized in Chapter 2. While the majority of the remote sensing systems may have been initially developed for military reconnaissance applications, the systems are now also heavily used for monitoring the Earth's natural resources.

The remote sensing data-collection and analysis procedures used for Earth resource applications are often implemented in a systematic fashion that can be termed the **remote sensing process**. The procedures in the process are summarized here and in Figure 1-5:

- The hypothesis to be tested is defined using a specific type of logic (e.g., inductive, deductive) and an appropriate processing model (e.g., deterministic, stochastic).
- *In situ* and collateral information necessary to calibrate the remote sensor data and/or judge its geometric, radiometric, and thematic characteristics are collected.
- Remote sensor data are collected passively (e.g., digital cameras) or actively (e.g., RADAR, LiDAR) using analog or digital remote sensing instruments, ideally at the same time as the *in situ* data.
- *In situ* and remotely sensed data are processed using a variety of techniques, including: a) analog image processing, b) digital image processing, c) modeling, and d) *n*-dimensional visualization.
- Metadata, processing lineage, and the accuracy of the information are provided and the results communicated using images, graphs, statistical tables, GIS databases, Spatial Decision Support Systems (SDSS), etc.

It is useful to review the characteristics of these remote sensing process procedures.

Statement of the Problem

Sometimes the general public and even children look at aerial photography or other remote sensor data and extract useful information. They do this without a formal hypothesis in mind. Often, however, they interpret the imagery incorrectly because they do not understand the nature of the remote sensing system used to collect the data or appreciate the vertical or oblique perspective of the terrain recorded in the imagery.

Scientists who use remote sensing, on the other hand, are usually trained in the scientific method—a way of thinking about problems and solving them. They use a formal plan that typically has at least five elements: 1) stating the problem, 2) forming the research hypothesis (i.e., a possible explanation), 3) observing and experimenting, 4) interpreting data, and 5) drawing conclusions. It is not necessary to follow this formal plan exactly.

FIGURE 1–5 Scientists generally use the remote sensing process to extract information from remotely-sensed images.

The scientific method is normally used in conjunction with environmental models that are based on two primary types of logic:

- deductive logic, and
- inductive logic.

Models based on deductive and/or inductive logic can be further subdivided according to whether they are processed deterministically or stochastically. Some scientists extract new thematic information directly from remotely sensed imagery without ever explicitly using inductive or deductive logic. They are just interested in extracting information from the imagery using appropriate methods and technology. This technological approach is not as rigorous, but it is common in applied remote sensing. The approach can also generate new knowledge.

Remote sensing is used in both scientific (inductive and deductive) and technological approaches to obtain knowledge. There is discussion as to how the different types of logic used in the remote sensing process yield new scientific knowledge (e.g., Fussell et al., 1986; Curran, 1987; Fisher and Lindenberg, 1989; Dobson, 1993; Skidmore, 2002; Wulder and Coops, 2014).

Identification of *In situ* and Remote Sensing Data Requirements

If a hypothesis is formulated using inductive and/or deductive logic, a list of variables or observations are identified that will be used during the investigation. *In situ* observation and/or remote sensing may be used to collect information on the most important variables.

Scientists using remote sensing technology should be well trained in field and laboratory data-collection procedures. For example, if a scientist wants to map the surface temperature of a lake, it is usually necessary to collect accurate empirical *in situ* lake-temperature measurements at the same time the remote sensor data are collected. The *in situ* data may be used to: 1) calibrate

the remote sensor data, and/or 2) perform an unbiased accuracy assessment of the final results (Congalton and Green, 2009). Remote sensing textbooks provide some information on field and laboratory sampling techniques (e.g., Jensen, 2007). The *in situ* sampling procedures, however, are learned best through formal courses in the sciences (e.g., chemistry, biology, forestry, soils, hydrology, meteorology). It is also important to know how to collect accurate socioeconomic and demographic information in urban environments based on training in human geography, sociology, etc. (e.g., McCoy, 2005; Azar et al., 2013).

Most *in situ* data are now collected in conjunction with global positioning system (GPS) *x*, *y*, *z* data (Jensen and Jensen, 2013). Scientists should know how to collect the GPS data at each *in situ* data-collection station and turn it into useful geospatial information.

Collateral Data Requirements

Many times collateral data (often called ancillary data), such as digital elevation models, soil maps, geology maps, political boundary files, and block population statistics, are of value in the remote sensing process. Ideally, the geospatial collateral data reside in a GIS (Jensen and Jensen, 2013).

Remote Sensing Data Requirements

Once we have a list of variables, it is useful to determine which of them can be remotely sensed. Remote sensing can provide information on two different classes of variables: biophysical and hybrid.

Biophysical Variables: Certain biophysical variables can be measured directly by a remote sensing system. This means that the remotely sensed data can provide fundamental biological and/or physical (biophysical) information directly, without having to use other surrogate or ancillary data (Wulder and Coops, 2014). For example, a thermal infrared remote sensing system can record the apparent surface temperature of a rock outcrop or agricultural field by measuring the radiant energy exiting its surface. Similarly, it is possible to conduct remote sensing in a very specific region of the spectrum and identify the amount of water vapor in the atmosphere. It is also possible to measure soil moisture content directly using active and passive microwave remote sensing techniques. NASA's Moderate Resolution Imaging Spectrometer (MODIS) can be used to model Absorbed Photosynthetically Active Radiation (APAR) and LAI. The precise *x*, *y* location, and height or elevation (*z*) of an object can be extracted directly from stereoscopic aerial photography, overlapping satellite imagery (e.g., SPOT), LiDAR data, or IFSAR imagery.

Table 1-1 is a list of selected biophysical variables that can be remotely sensed and useful sensors to acquire the data. Characteristics of many of these remote sensing systems are discussed in Chapter 2. Great strides have been made in remotely sensing many of these biophysical variables. They are important to the national and international effort under way to model the global environment (e.g., NRC, 2012; Brewin et al., 2013).

Hybrid Variables: The second general group of variables that can be remotely sensed includes hybrid variables, created by systematically analyzing more than one biophysical variable. For example, by remotely sensing a plant's chlorophyll absorption characteristics, temperature, and moisture content, it might be possible to model these data to detect vegetation stress, a hybrid variable. The variety of hybrid variables is large; consequently, no attempt is made to identify them. It is important to point out, however, that nominal-scale land use and land cover are hybrid variables. For example, the land cover of a particular area on an image may be derived by evaluating several of the fundamental biophysical variables at one time (e.g., object location [*x*, *y*], height [*z*], tone and/or color, biomass, and perhaps temperature). So much attention has been placed on remotely sensing this hybrid nominal-scale variable that the interval- or ratio-scaled biophysical variables were largely neglected until the mid-1980s. Nominal-scale land-use and land-cover mapping are important capabilities of remote sensing technology and should not be minimized. Many social and physical scientists routinely use such data in their research. However, there is now a dramatic increase in the extraction of interval- and ratio-scaled biophysical data that are incorporated into quantitative models that can accept spatially distributed information.

Remote Sensing Data Collection

Remotely sensed data are collected using passive or active remote sensing systems. **Passive sensors** record electromagnetic radiation that is reflected or emitted from the terrain. For example, cameras and video recorders can be used to record visible and near-infrared energy reflected from the terrain. A multispectral scanner can be used to record the amount of thermal radiant flux exiting the terrain. **Active sensors,** such as microwave (RADAR), LiDAR, or SONAR, bathe the terrain in machine-made electromagnetic energy and then record the time-lapsed amount of radiant flux scattered back toward the sensor system.

Remote sensing systems collect analog (e.g., hard-copy aerial photography or video data) and/or digital data (e.g., a matrix [raster] of brightness values obtained using a scanner, linear array, or area array).

TABLE 1–1 Selected biophysical and hybrid variables and potential remote sensing systems used to obtain the information.

Biophysical Variables	Potential Remote Sensing Systems
x, y, z **Location and Geodetic Control** ***x, y*** **Location from Orthocorrected Imagery**	- Global Positioning Systems (GPS) - Analog and digital stereoscopic aerial photography, GeoEye-1, WorldView-2, SPOT 6 and 7, Landsat 7 and 8, ResourceSat, ERS-1 and -2, Moderate Resolution Imaging Spectrometer (MODIS), Pleiades, LiDAR, RADARSAT-1 and -2
z **Topography/Bathymetry** - Digital Elevation Model (DEM) - Digital Bathymetric Model (DBM)	- GPS, stereoscopic aerial photography, LiDAR, SPOT 6, RADARSAT, GeoEye-1, WorldView-2 and -3, Shuttle Radar Topography Mission (SRTM), IFSAR - SONAR, bathymetric LiDAR, stereoscopic aerial photography
Vegetation - Pigments (e.g., chlorophyll *a* and *b*) - Canopy structure and height - Biomass derived from vegetation indices - LAI - Absorbed photosynthetically active radiation - Evapotranspiration	- Color aerial photography, Landsat 8, GeoEye-1, WorldView-2 and -3, Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), MODIS, Pleiades, airborne hyperspectral systems (e.g., AVIRIS, HyMap, CASI) - Stereoscopic aerial photography, LiDAR, RADARSAT, IFSAR - Color-infrared (CIR) aerial photography, Landsat 8, GeoEye-1, WorldView-2, Advanced Very High Resolution Radiometer (AVHRR), Multiangle Imaging Spectroradiometer (MISR), Pleiades, satellite (EO-1 Hyperion) and airborne hyperspectral systems (e.g., AVIRIS, HyMap, CASI-1500)
Surface Temperature (land, water, atmosphere)	- ASTER, AVHRR, GOES, Hyperion, Landsat 8, MODIS, airborne thermal infrared
Soil and Rocks - Moisture - Mineral composition - Taxonomy - Hydrothermal alteration	- ASTER, passive microwave (SSM/1), RADARSAT-2, MISR, ALMAZ, Landsat 8, ERS-1 and -2 - ASTER, MODIS, airborne and satellite hyperspectral systems - Color and CIR aerial photography, airborne hyperspectral systems - Landsat 8, ASTER, MODIS, airborne and satellite hyperspectral systems
Surface Roughness	- Aerial photography, RADARSAT-1 and -2, IKONOS-2, WorldView-2, ASTER
Atmosphere - Aerosols (e.g., optical depth) - Clouds (e.g., fraction, optical thickness) - Precipitation - Water vapor (precipitable water) - Ozone	- NPP, MISR, GOES, AVHRR, MODIS, CERES, MOPITT, MERIS - NPP, GOES, AVHRR, MODIS, MISR, CERES, MOPITT, UARS, MERIS - Tropical Rainfall Measurement Mission (TRMM), GOES, AVHRR, SSM/1, MERIS - NPP, GOES, MODIS, MERIS - NPP, MODIS
Water - Color - Surface hydrology - Suspended minerals - Chlorophyll/gelbstoffe - Dissolved organic matter	- Color and CIR aerial photography, Landsat 8, SPOT 6, GeoEye-1, WorldView-2, Pleiades, ASTER, MODIS, airborne and satellite hyperspectral systems, AVHRR, NPP, GOES, bathymetric LiDAR, MISR, CERES, TOPEX/POSEIDON, MERIS
Snow and Sea Ice - Extent and characteristics	- Color and CIR aerial photography, NPP, AVHRR, GOES, Landsat 8, SPOT 6, GeoEye-1, WorldView-2, Pleiades, ASTER, MODIS, MERIS, ERS-1 and -2, RADARSAT-1 and -2
Volcanic Effects - Temperature, gases	- ASTER, Landsat 8, MISR, Hyperion, MODIS, thermal hyperspectral systems
BRDF (bidirectional reflectance distribution function)	- MISR, MODIS, CERES
Selected Hybrid Variables	**Potential Remote Sensing Systems**
Land Use - Commercial, residential, transportation, etc. - Cadastral (property) **Land Cover** - Agriculture, forest, urban, etc.	- High spatial resolution panchromatic, color and /or CIR stereoscopic aerial photography, high spatial resolution satellite imagery (<1 x 1 m: GeoEye-1, WorldView-3), SPOT 6 and 7, LiDAR, high spatial resolution hyperspectral systems - Color and CIR aerial photography, Landsat 8, SPOT 6 and 7, ASTER, AVHRR, RADARSAT, GeoEye-1, WorldView-2 and -3, Pleiades, LiDAR, IFSAR, MODIS, MISR, MERIS, airborne and satellite hyperspectral systems
Vegetation - stress - composition/type	- Color and CIR photography, Landsat 8, GeoEye-1, WorldView-2, AVHRR, SeaWiFS, MISR, MODIS, ASTER, MERIS, airborne and satellite hyperspectral systems

The amount of electromagnetic radiance, L (watts m^{-2} sr^{-1}; watts per meter squared per steradian), recorded within the IFOV of an optical remote sensing system (e.g., a pixel in a digital image), is a function of:

$$L = f(\lambda, s_{x,y,z}, t, \beta, \theta, P, \Omega) \tag{1.1}$$

where

λ = wavelength (spectral response measured in various bands or at specific frequencies). Wavelength (λ) and frequency (υ) may be used interchangeably based on their relationship with the speed of light (c) where $c = \lambda \times \upsilon$.

$s_{x,y,z}$ = x, y, z location of the pixel and its size (x, y);

t = temporal information, i.e., when, how long, and how often the data were acquired;

β = IFOV;

θ = set of angles that describe the geometric relationships between the radiation source (e.g., the Sun), the terrain target of interest (e.g., a wheat field), and the remote sensing system;

P = polarization of back-scattered energy recorded by the sensor; and

Ω = radiometric resolution (precision) at which the data (e.g., reflected, emitted, or back-scattered radiation) are recorded by the remote sensing system.

It is useful to review characteristics of the parameters associated with Equation 1.1 and how they influence the nature of the remote sensing data collected.

Spectral Information and Resolution

Most remote sensing investigations are based on developing a deterministic relationship (i.e., a model) between the amount of electromagnetic energy reflected, emitted, or back-scattered in specific bands or frequencies and the chemical, biological, and physical characteristics of the phenomena under investigation (e.g., a wheat field canopy). **Spectral resolution** is the number and dimension (size) of wavelength intervals (referred to as bands or channels) in the electromagnetic spectrum to which a remote sensing instrument is sensitive.

Multispectral remote sensing systems record energy in multiple bands of the electromagnetic spectrum. For example, in the 1970s and early 1980s, the Landsat Multispectral Scanner (MSS) recorded remotely sensed data of much of the Earth that is still of significant value for change detection studies. The bandwidths of the four MSS-bands are displayed in Figure 1-6a (band 1 = 500 – 600 nm; band 2 = 600 – 700 nm; band 3 = 700 – 800 nm; and band 4 = 800 – 1,100 nm). The nominal size of a band may be large (i.e., coarse), as with the Landsat MSS near-infrared band 4 (800 – 1,100 nm) or relatively smaller (i.e., finer), as with the Landsat MSS band 3 (700 – 800 nm).

The four multispectral bandwidths associated with a typical digital frame camera are also shown in Figure 1-6a. The camera's detectors record information in four regions of the spectrum (band 1 = 450 – 515 nm; band 2 = 525 – 605 nm; band 3 = 640 – 690 nm; and band 4 = 750 – 900 nm). Note that there are gaps between the spectral sensitivities of the detectors. Individual band images are shown in Figure 1-6c. Natural and color-infrared color composites created using the individual bands are shown in Figure 1-6d-e.

In reality, the spectral sensitivity of an individual band does not cut off as neatly as shown in Figure 1-6a. Instead, detectors are sensitive over a region in the spectrum with a maximum intensity at approximately the middle of the bandwidth. Therefore, the values used to describe the bandwidth of an individual band are usually derived using the logic shown in Figure 1-6b where the intensity and shape of the Gaussian curve is investigated and the **Full Width at Half Maximum** (FWHM) is determined. In this hypothetical example, the FWHM of band 2 (green) is from 525 to 605 nm. Therefore, we can reasonably say that this is the spectral bandwidth of this particular band, even though we know it has some sensitivity outside of this bandwidth. The band center is at 565 nm. Sometimes scientists use band center information in their investigations, especially when analyzing hyperspectral imagery.

A **hyperspectral remote sensing** instrument typically acquires data in hundreds of spectral bands. For example, a hyperspectral image of Sullivan's Island, SC, collected by NASA's Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) is shown in Figure 1-7a. AVIRIS collects data in 224 bands in the region from 400 – 2,500 nm spaced just 10 nm apart based on the FWHM criteria (NASA AVIRIS, 2014). The bandwidths of the nine Landsat 8 sensor system are shown in Figure 1-7b for comparison (Irons et al., 2012). Note how AVIRIS obtains hundreds of measurements per pixel over the 400 – 2,500 nm region. Some scientists like to acquire the complete spectrum of information from 400 – 2,500 nm. Others prefer to work on a relatively small subset of carefully placed bands such as those associated with Landsat 8 to extract the desired information. **Ultraspectral remote sensing** involves data collection in many hundreds of bands.

Certain regions or spectral bands of the electromagnetic spectrum are optimal for obtaining information on

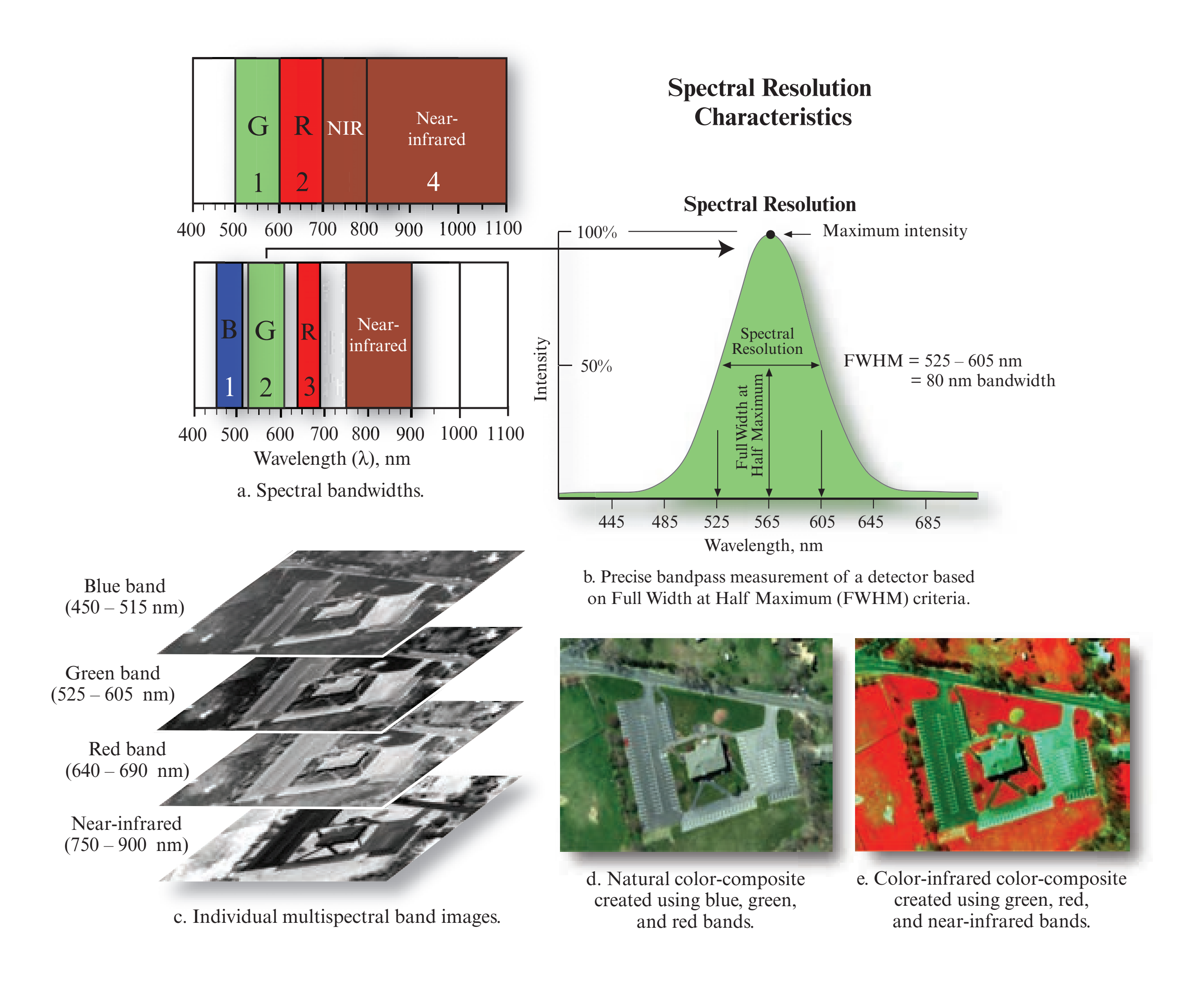


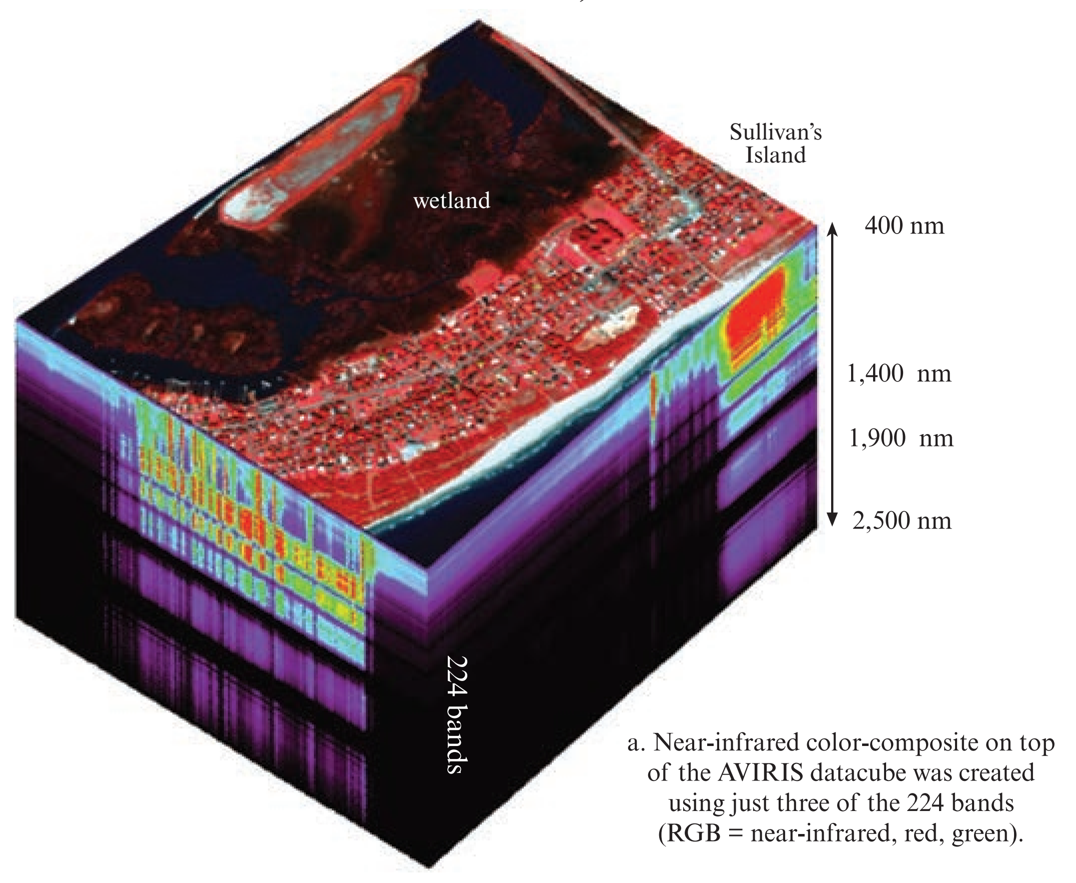
FIGURE 1–6 a) The relatively coarse spectral bandwidths of the four Landsat Multispectral Scanner (MSS) bands compared with the bandwidths of a typical digital frame camera. b) The true spectral bandwidth is the width of the Gaussian-shaped spectral profile based on Full Width at Half Maximum (FWHM) intensity. For example, at FWHM the green band has a spectral range of 80 nm between 525 and 605 nm. c) Examples of 1 × 1 ft. spatial resolution individual digital frame camera images. d) Natural color composite created using the blue, green, and red bands. e) Color-infrared color-composite created using the green, red, and near-infrared bands.

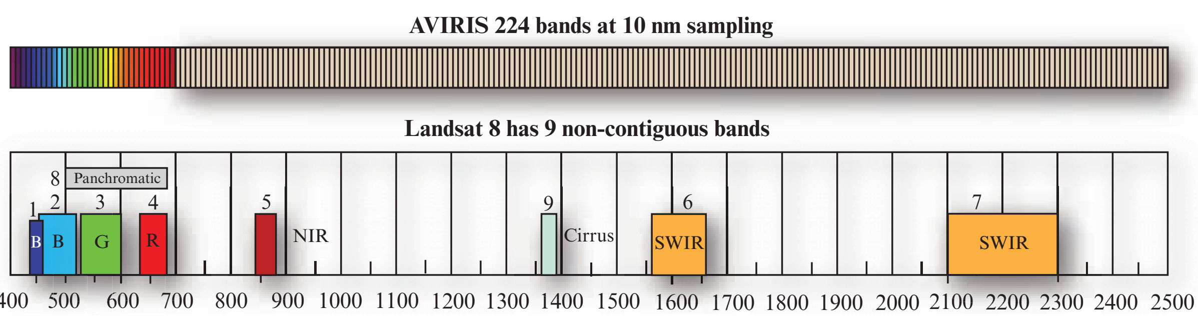
biophysical parameters. The bands are normally selected to maximize the contrast between the object of interest and its background (i.e., object-to-background contrast). Careful selection of the spectral bands might improve the probability that the desired information will be extracted from the remote sensor data.

Spatial Information and Resolution

Most remote sensing studies record the spatial attributes of objects on the terrain. For example, each silver halide crystal in an analog aerial photograph and each picture element in a digital remote sensor image is located at a specific location in the image and associated with specific *x*, *y* coordinates on the ground. Once geographically referenced to a standard map projection (Chapter 7), the spatial information associated with each silver halide crystal or pixel is of significant value because it allows the remote sensing–derived information to be used with other spatial data in a GIS or spatial decision support system (Jensen et al., 2002).

There is a general relationship between the size of an object or area to be identified and the spatial resolution of the remote sensing system. **Spatial resolution** is a measure of the smallest angular or linear separation between two objects that can be resolved by the remote sensing system. The spatial resolution of aerial photography may be measured by 1) placing calibrated, parallel black and white lines on tarps that are placed in the field, 2) obtaining aerial photography of the study area, and 3) computing the number of resolvable line pairs per millimeter in the photography. It is also possible to determine the spatial resolution of imagery by





b. Comparison of 224 AVIRIS spectral bands at 10 nm sampling with the sensitivity of the nine Landsat 8 non-contiguous bands in the region from 400 to 2,500 nm.

FIGURE 1–7 a) An AVIRIS hyperspectral datacube of Sullivan's Island, SC. The image on top is a color composite of just three of the 224 available bands (RGB = near-infrared, red, green). b) A comparison of the sensitivity of the 224 AVIRIS bands with the location of the nine Landsat 8 non-contiguous bands in the region from 400 to 2,500 nm. Source of data: NASA.

computing its modulation transfer function, which is beyond the scope of this text.

Many satellite remote sensing systems use optics that have a constant instantaneous-field-of-view (IFOV) (Figure 1-2). Therefore, a sensor system's nominal spatial resolution is defined as the dimension in meters (or feet) of the ground-projected IFOV where the diameter of the circle (D) on the ground is a function of the instantaneous-field-of-view (β) times the altitude (H) of the sensor above ground level (AGL) (Figure 1-2):

$$D = \beta \times H. \tag{1.2}$$

Pixels are normally represented on computer screens and in hard-copy images as rectangles with length and width. Therefore, we typically describe a sensor system's nominal spatial resolution as being 10 × 10 m or 30 × 30 m. For example, DigitalGlobe's WorldView-2

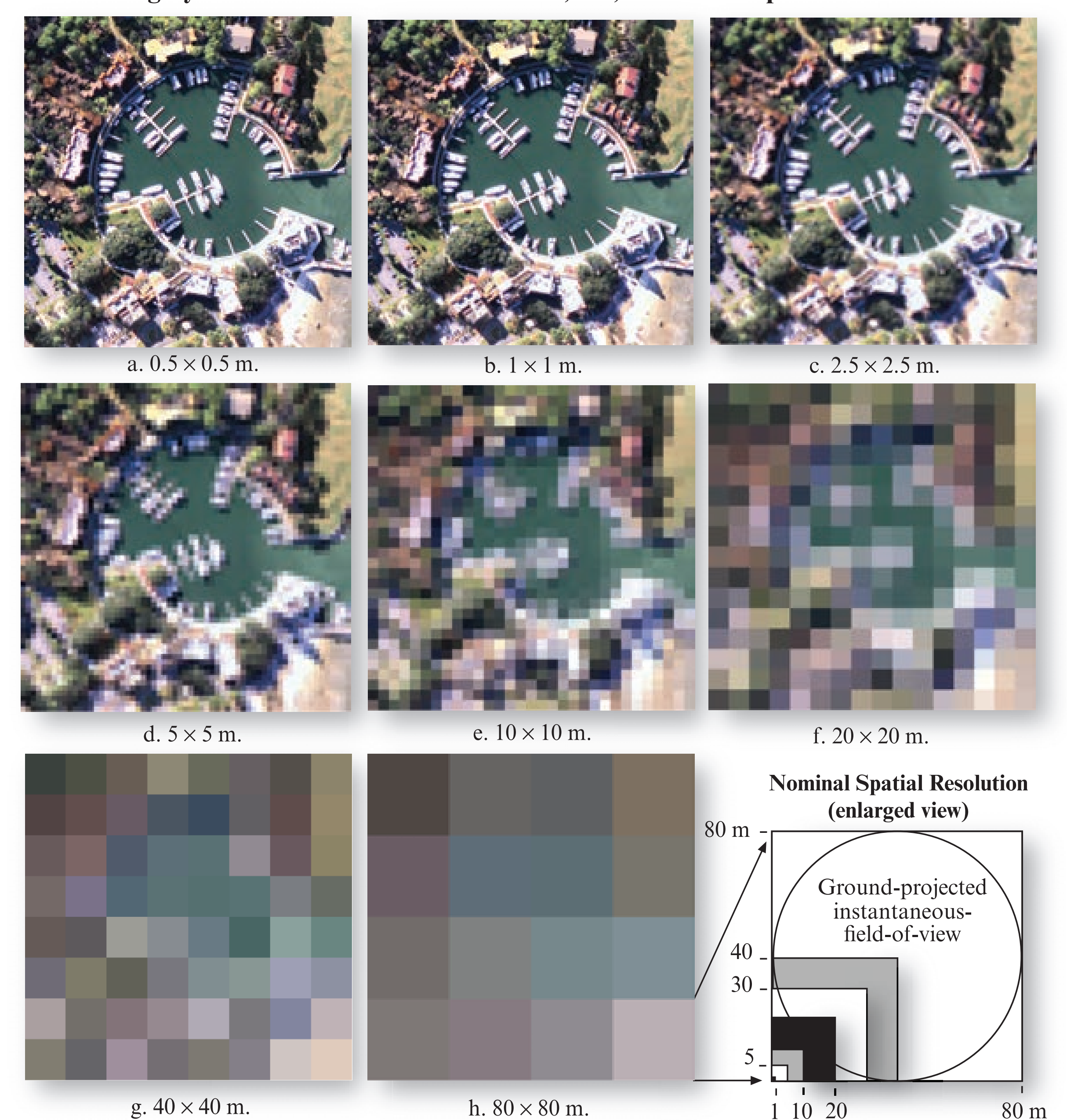


FIGURE 1–8 The original Harbor Town, SC, image was collected at a nominal spatial resolution of 0.3 × 0.3 m (approximately 1 × 1 ft.) using a digital frame camera. The original imagery was resampled to derive the imagery with the simulated spatial resolutions shown.

has a nominal spatial resolution of 0.46 × 0.46 cm for its panchromatic band and 1.85 × 1.85 m for the four multispectral bands. The Landsat 7 Enhanced Thematic Mapper Plus (ETM+) has a nominal spatial resolution of 15 × 15 m for its panchromatic band and 30 × 30 m for six of its multispectral bands. Generally, the smaller the nominal spatial resolution, the greater the spatial resolving power of the remote sensing system.

Figure 1-8 depicts digital camera imagery of an area in Hilton Head, SC, at resolutions ranging from 0.5 × 0.5 m to 80 × 80 m. Note that there does not appear to be a significant difference in the interpretability of 0.5 × 0.5 m data and 1 × 1 m data. However, the urban spatial information content decreases rapidly when using 5 × 5 m imagery and is practically useless for urban analysis at spatial resolutions greater than 10 × 10 m. This is the reason historical Landsat MSS data (79 × 79 m) are of little value for most urban applications (Jensen and Cowen, 1999).

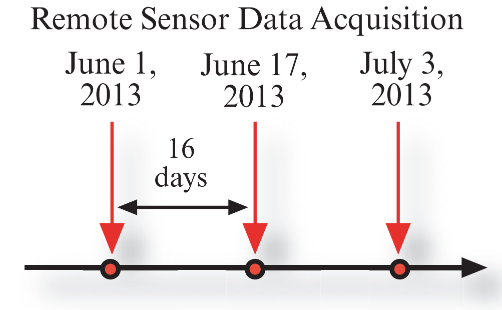


FIGURE 1–9 The temporal resolution of a remote sensing system refers to how often it records imagery of a particular area. This example depicts the systematic collection of remote sensor data every 16 days, presumably at approximately the same time of day. NASA Landsat Thematic Mappers 4 and 5 had 16-day revisit cycles. The NOAA Geostationary Operational Environmental Satellites (GOES) collect new images every half-hour which are especially useful for monitoring storm events in near real-time.

A useful heuristic rule of thumb is that in order to detect a feature, the nominal spatial resolution of the remote sensing system should be less than one-half the size of the feature measured in its smallest dimension. For example, if we want to identify the location of all oak trees in a park, the minimum acceptable spatial resolution would be approximately one-half the diameter of the smallest oak tree's crown. Even this spatial resolution, however, will not guarantee success if there is no difference between the spectral response of the oak tree (the object) and the soil or grass surrounding it (i.e., its background).

Some sensor systems, such as LiDAR, do not completely "map" the terrain surface. Rather, the surface is "sampled" using laser pulses sent from the aircraft at some nominal time interval (Renslow, 2012). The ground-projected laser pulse may be very small (e.g., 10–15 cm in diameter) with samples located approximately every 1–6 m on the ground. Spatial resolution would appropriately describe the ground-projected laser pulse (e.g., 15 cm) but sampling density (i.e., number of points per unit area) describes the frequency of ground observations (Hodgson et al., 2005).

Because we have spatial information about the location of each pixel (x, y) in the image matrix, it is also possible to examine the spatial relationship between a pixel and its neighbors. Therefore, the amount of spatial autocorrelation and other spatial geostatistical measurements can be determined based on the spatial information inherent in the imagery.

Temporal Information and Resolution

One of the valuable things about remote sensing science is that it obtains a record of Earth landscapes at a unique moment in time. Multiple records of the same landscape obtained through time can be used to identify processes at work and to make predictions.

The **temporal resolution** of a remote sensing system generally refers to how often and when the sensor records imagery of a particular area. The temporal resolution of the sensor system shown in Figure 1-9 is every 16 days. Ideally, the sensor obtains data repetitively to capture unique discriminating characteristics of the object under investigation. For example, agricultural crops have unique developmental **phenological cycles** in each geographic region. To measure specific agricultural variables, it is necessary to acquire remotely sensed data at critical dates in the phenological cycle. Analysis of multiple-date imagery provides information on how the variables are changing through time. Change information provides insight into processes influencing the development of the crop. Fortunately, several satellite sensor systems such as SPOT 4 and 5, GeoEye-1, ImageSat and WorldView-2 are pointable, meaning that they can acquire imagery off-nadir. Nadir is the point directly below the spacecraft. This dramatically increases the probability that imagery will be obtained during a growing season or during an emergency. However, off-nadir oblique viewing also introduces bidirectional reflectance distribution function (BRDF) issues, discussed in the next section.

There are often trade-offs associated with the various resolutions that must be made when collecting remote sensing data (Figure 1-10). Generally, the higher the temporal resolution requirement (e.g., monitoring hurricanes every half-hour), the lower the spatial resolution requirement (e.g., the NOAA GOES weather satellite records images with 4 × 4 to 8 × 8 km pixels). Conversely, the higher the spatial resolution requirement (e.g., monitoring urban land-use with 1 × 1 m data), the lower the temporal resolution requirement (e.g., every 1 to 10 years). For example, Figure 1-11 documents single-family residential land-use development for an area near White Rock, SC, using high spatial resolution (1 × 1 ft.) digital aerial photography collected in 2004, 2007, and 2009. Some applications such as crop type or yield estimation might require relatively high temporal resolution data (e.g., multiple images obtained during a growing season) and moderate spatial resolution data (e.g., 250 × 250 m pixels provided by NASA's MODIS sensor). Emergency response applications may require very high spatial and temporal resolution data collection that generates tremendous amounts of data (e.g., every 5 hours at 0.5 × 0.5 m spatial resolution).

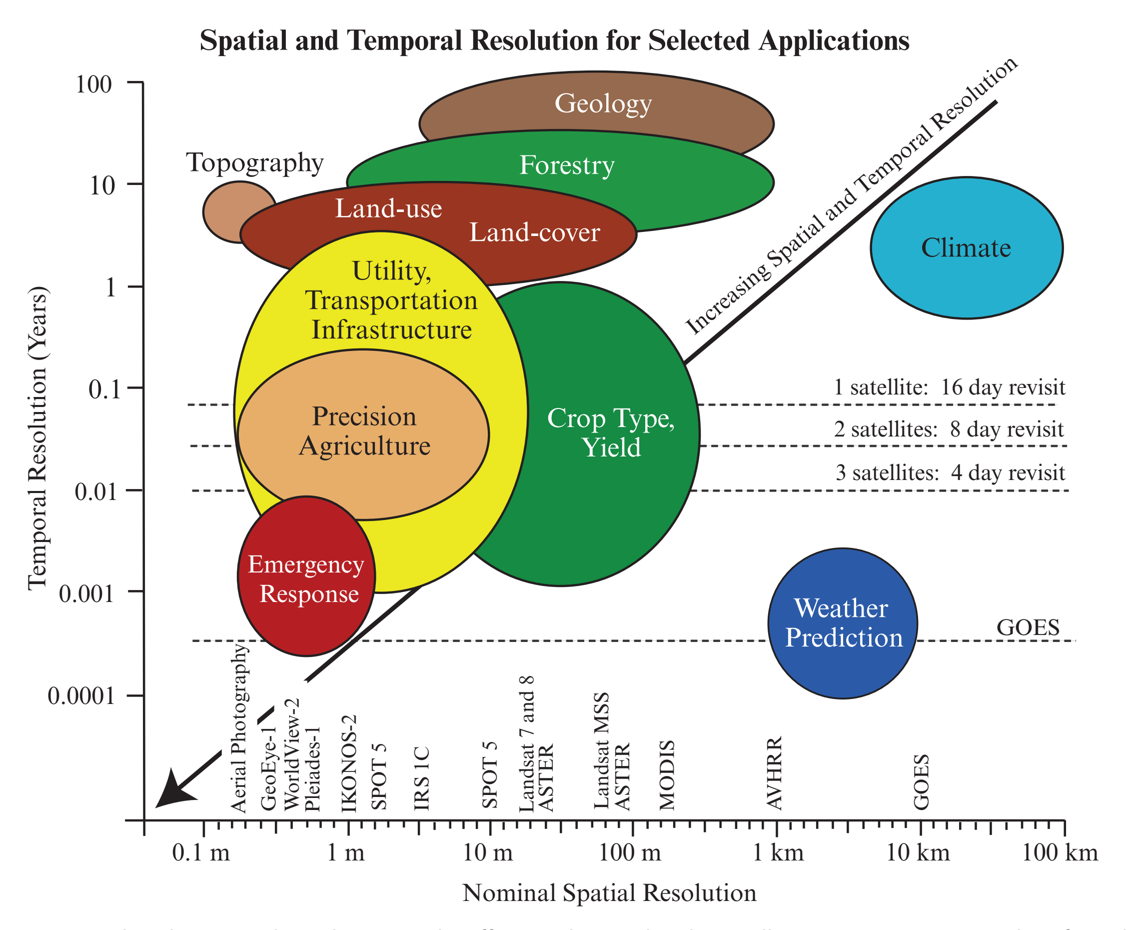


FIGURE 1–10 Spatial and temporal resolution trade-offs must be made when collecting remote sensor data for selected applications. For example, applications such as land use mapping generally require high spatial resolution imagery (e.g., 1 to 5 m) at relatively low temporal resolution (e.g., 1–10 years). Conversely, for weather prediction we are generally content with lower spatial resolution imagery (e.g., 5 × 5 km) if it can be collected frequently (e.g., every half-hour).

Another aspect of temporal information is how many observations are recorded from a single pulse of energy that is directed at the Earth by an active sensor such as LiDAR. For example, most LiDAR sensors emit one pulse of laser energy and record multiple responses from this pulse. Measuring the time differences between multiple responses allows for the determination of object heights and terrain structure.

Radiometric Information and Resolution

Some remote sensing systems record the reflected, emitted, or back-scattered electromagnetic radiation with more precision than other sensing systems. This is analogous to making a measurement with a ruler. If you want precisely to measure the length of an object, would you rather use a ruler with 16 or 1,024 subdivisions on it?

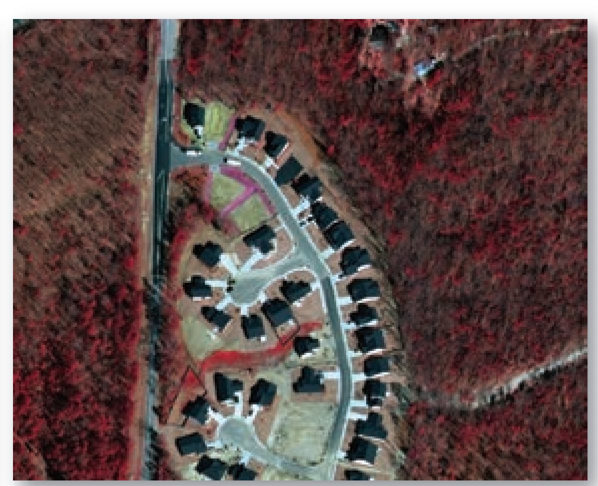
Radiometric resolution is defined as the sensitivity of a remote sensing detector to differences in signal strength as it records the radiant flux reflected, emitted, or back-scattered from the terrain. It defines the number of just discriminable signal levels. Therefore, radiometric resolution can have a significant impact on our ability to measure the properties of scene objects. The Landsat 1 Multispectral Scanner launched in 1972 recorded reflected energy with a precision of 6-bits (values ranging from 0 to 63). Landsat 4 and 5 Thematic Mapper sensors launched in 1982 and 1984, respectively, recorded data in 8-bits (values from 0 to 255) (Figure 1-12). Thus, the Landsat TM sensors had improved radiometric resolution (sensitivity) when compared with the original Landsat MSS. GeoEye-1 and WorldView-2 sensors record information in 11-bits (values from 0 to 2,047). Several new sensor systems have 12-bit radiometric resolution (values ranging from 0 to 4,095). Radiometric resolution is sometimes referred to as the **level of quantization**. High radiometric resolution generally increases the probability that phenomena will be remotely sensed more accurately.

Polarization Information

The polarization characteristics of electromagnetic energy recorded by a remote sensing system are an important variable that can be used in many Earth resource investigations. Sunlight is polarized weakly.



a. 2004 color-infrared 1 × 1 ft. color composite.



b. 2007 color-infrared 1 × 1 ft. color composite.



c. 2009 natural-color 1 × 1 ft. color composite.

Multiple-date Digital Frame Camera Aerial Photography of an Area Near White Rock, SC

FIGURE 1–11 Digital frame camera aerial photography of residential development near White Rock, SC, collected in 2004, 2007, and 2009. Aerial photography used with permission of Richland County GIS Division.

However, when sunlight strikes a nonmetal object (e.g., grass, forest, or concrete) it becomes depolarized and the incident energy is scattered differentially. Generally, the more smooth the surface, the greater the polarization. It is possible to use polarizing filters on passive remote sensing systems (e.g., aerial cameras) to record polarized light at various angles. It is also possible to selectively send and receive polarized energy using active remote sensing systems such as RADAR (e.g., horizontal send, vertical receive—HV; vertical send, horizontal receive—VH; vertical send, vertical receive—VV; horizontal send, horizontal receive—HH). Multiple-polarized RADAR imagery is an especially useful application of polarized energy.

Angular Information

Remote sensing systems record specific angular characteristics associated with each exposed silver halide

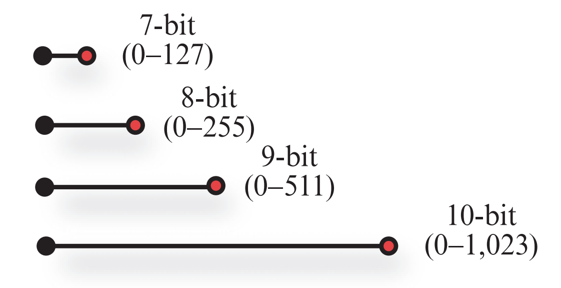


FIGURE 1–12 The radiometric resolution of a remote sensing system is the sensitivity of its detectors to differences in signal strength as they record the radiant flux reflected, emitted, or back-scattered from the terrain. The energy is normally quantized during an analog-to-digital conversion process to 8-, 9-, 10-bits or more.

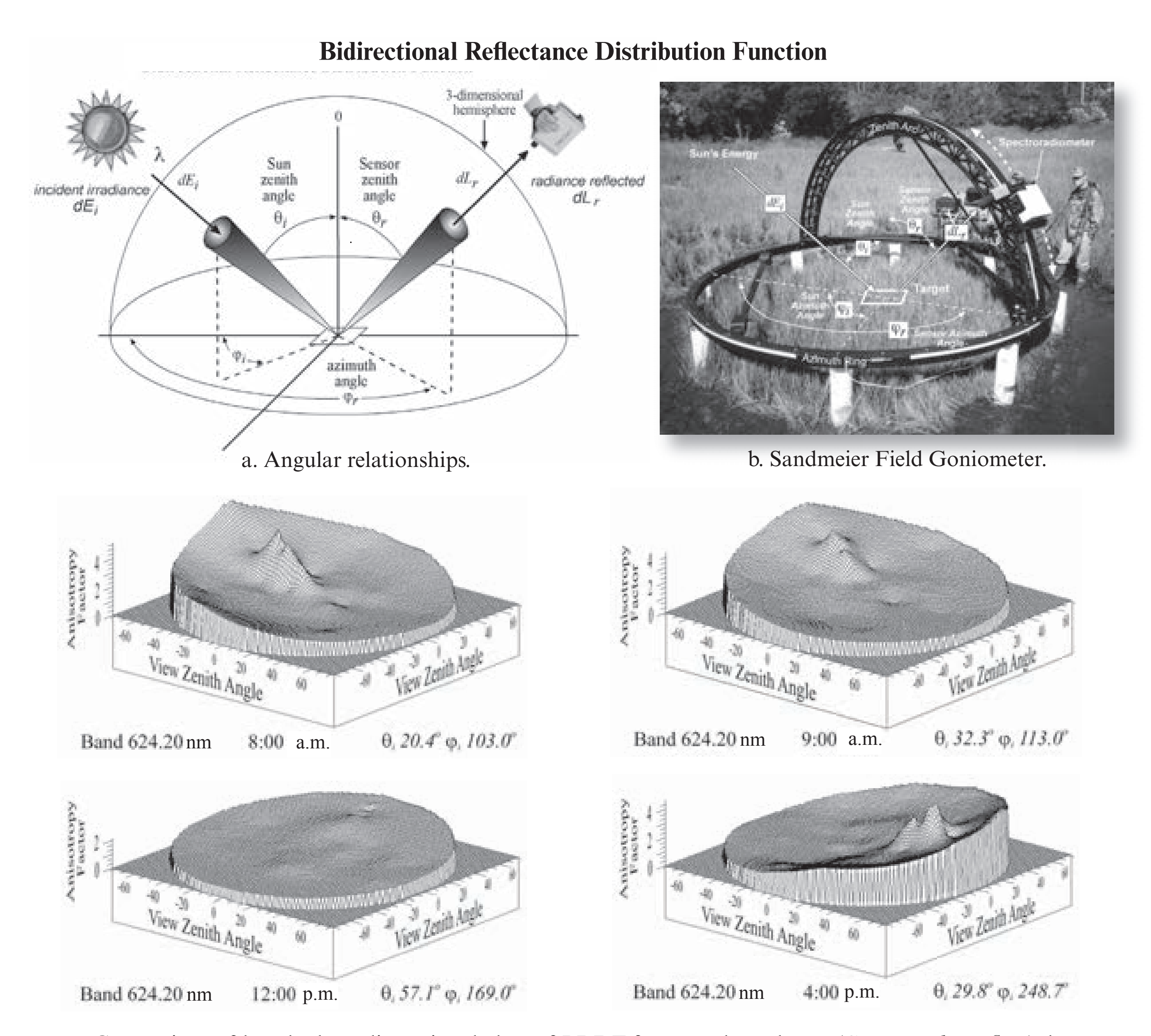


FIGURE 1–13 a) The concepts and parameters of the bidirectional reflectance distribution function (BRDF). A target is bathed in irradiance (dE_i) from a specific Sun zenith and azimuth angle, and the sensor records the radiance (dL_r) exiting the target of interest at a specific azimuth and zenith angle. b) The Sandmeier Field Goniometer collecting smooth cordgrass (*Spartina alterniflora*) BRDF measurements at North Inlet, SC. Spectral measurements are made at Sun zenith angle of θ_i and Sun azimuth angle of φ_i and a sensor zenith angle of view of θ_r and sensor azimuth angle of φ_r. A GER 3700 spectroradiometer, attached to the moving sled mounted on the zenith arc, recorded the amount of radiance leaving the target in 704 bands at 76 angles. c) Hourly three-dimensional plots of BRDF data. Diagrams and photograph adapted from Schill et al., 2004.

crystal or detector element. The angular characteristics are a function of (Figure 1-13a):

- the location in a three-dimensional sphere of the illumination source (e.g., the Sun for a passive system or the sensor itself in the case of RADAR, LiDAR, and SONAR) and its associated azimuth and zenith angles,
- the orientation of the terrain facet (pixel) or terrain cover (e.g., vegetation) under investigation, and
- the location of the suborbital or orbital remote sensing system and its associated azimuth and zenith angles.

There is always an angle of incidence associated with the incoming energy that illuminates the terrain and an angle of exitance from the terrain to the sensor system. This **bidirectional** nature of remote sensing data collection influences both the spectral and the polarization

characteristics of the at-sensor radiance, *L*, recorded by the remote sensing system.

A **goniometer** can be used to measure the changes in at-sensor radiance, *L*, caused by changing the position of the sensor and/or the source of the illumination (e.g., the Sun) (Figure 1-13b). For example, Figure 1-13c presents three-dimensional plots of smooth cordgrass (*Spartina alterniflora*) wetland BRDF data collected at 8 a.m., 9 a.m., 12 p.m., and 4 p.m. on March 21, 2000, for band 624.20 nm at the Baruch Marine Lab near Georgetown, SC. The only thing that changed between observations was the Sun's azimuth and zenith angles. The azimuth and zenith angles of the spectroradiometer were held constant while viewing the smooth cordgrass. Ideally, the BRDF plots would be identical, suggesting that it does not matter what time of day we collect the remote sensor data because the spectral reflectance characteristics from the smooth cordgrass remain constant. It is clear that this is not the case and that the time of day influences the spectral response. The Multiangle Imaging Spectrometer (MISR) onboard the *Terra* satellite was designed to investigate the BRDF phenomena. Research continues on how to incorporate the BRDF information into the digital image processing system to improve our understanding of what is recorded in the remotely sensed imagery (e.g., Sandmeier, 2000; Schill et al., 2004).

Angular information is central to the use of remote sensor data in photogrammetric applications. Stereoscopic image analysis is based on the assumption that an object on the terrain is remotely sensed from two angles. Viewing the same terrain from two vantage points introduces stereoscopic parallax, which is the foundation for all stereoscopic photogrammetric and radargrammetric analysis (Wolf et al., 2013).

Suborbital (Airborne) Remote Sensing Systems

High-quality photogrammetric cameras mounted onboard aircraft continue to provide aerial photography for many Earth resource applications. For example, photogrammetric engineering firms such as Pictometry International, Inc. and Sanborn Map Company provide both vertical and oblique digital aerial photography for many of the counties in the United States. These high spatial resolution datasets are crucial for maintaining local infrastructure including hydrologic networks, transportation networks, utilities, land use planning, zoning, tax mapping, impervious surface mapping, etc.

In addition, remote sensing systems are routinely mounted on aircraft to provide high spatial and spectral resolution remotely sensed data. Examples include hyperspectral sensors such as NASA's Airborne Visible-Infrared Imaging Spectrometer (AVIRIS), the Canadian Airborne Imaging Spectrometer (CASI-1500), and the Australian HyMap hyperspectral system discussed in Chapter 2 and Chapter 11. These sensors can collect data on demand when disaster strikes if cloud-cover conditions permit. For example, NASA collected hundreds of flightlines of AVIRIS data to monitor the *Deepwater Horizon* oil spill in the Gulf of Mexico in 2010. Unfortunately, suborbital remote sensor data are expensive to acquire per km^2. Also, atmospheric turbulence can cause the data to have severe geometric distortions that can be difficult to correct.

Satellite Remote Sensing Systems

Remote sensing systems onboard satellites provide high-quality, relatively inexpensive data per km^2. The United States has progressed from multispectral scanning systems (Landsat MSS launched in 1972) to more advanced scanning systems (Landsat 8 was launched on February 11, 2013). The Land Remote Sensing Policy Act of 1992 specified the future of satellite land remote sensing programs in the United States (Jensen, 1992). Unfortunately, Landsat 6 with its Enhanced Thematic Mapper did not achieve orbit when launched on October 5, 1993. Landsat 7 was launched on April 15, 1999, to relieve the United States' land remote sensing data gap. Unfortunately, it now has serious scan-line corrector problems.

The NASA Earth Observing-1 (EO-1) mission launched on November 21, 2000, validated a multispectral instrument (the Advanced Land Imager—ALI) that is a significant improvement over the Landsat 7 ETM^+ instrument; validated a hyperspectral land imaging instrument (Hyperion) and the unique science that can be performed with hyperspectral data; and validated the ability of a low-spatial/high-spectral resolution imager that can correct systematic errors in the apparent surface reflectance caused by atmospheric effects, primarily water vapor. The breakthrough spacecraft bus technologies that were validated were made available for Landsat 8 launched in 2013.

Meanwhile, numerous other countries launched high and moderate and high spatial resolution panchromatic and multispectral remote sensing systems during the first decade of the twenty-first century (e.g., France [SPOT 5 and 6, Pleiades], Korea [KompsSat-2 and -3], and India [CartoSat-2, ResourceSat-2]). These Earth resource satellites are described in Chapter 2.

Global Climate Change and Satellite Remote Sensing: The Earth's physical climate subsystem is sensitive to changes in the Earth's radiation balance. The International Geosphere–Biosphere Program (IGBP) and the U.S. Global Change Research Program (USGCRP) both call for scientific research to describe and understand the interactive physical, chemical, and

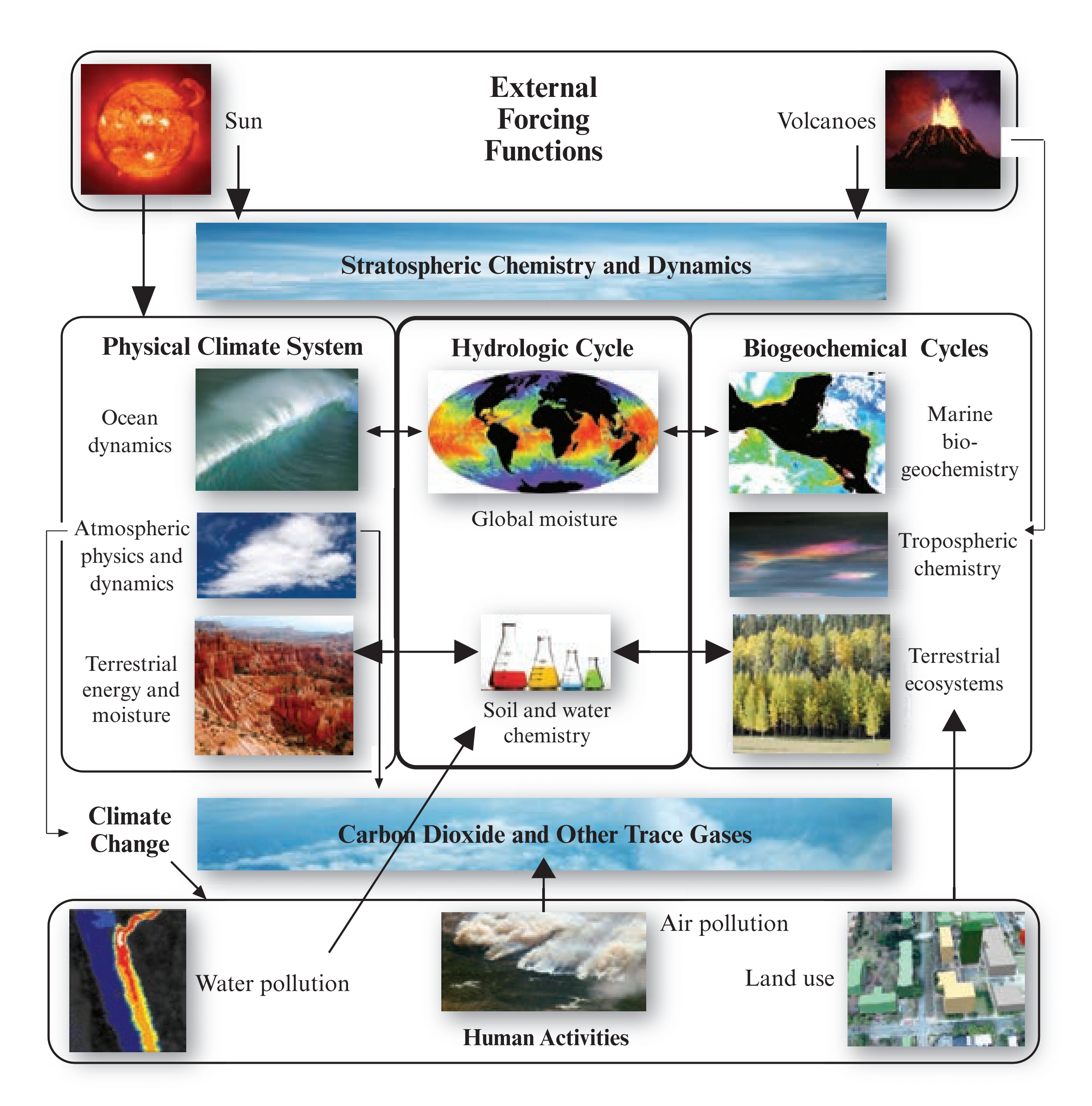


FIGURE 1–14 The Earth system can be subdivided into two subsystems—the physical climate system and biogeochemical cycles—that are linked by the global hydrologic cycle. Significant changes in the external forcing functions and human activities have an impact on the physical climate system, biogeochemical cycles, and the global hydrologic cycle. Examination of these subsystems and their linkages defines the critical questions that the NASA Earth Science division is attempting to answer (concept updated from Asrar and Dozier, 1994; images courtesy of NASA, the U.S. Geological Survey, and the author).

biological processes that regulate the total Earth system (IGBP, 2014; USGCRP, 2014). Space-based remote sensing is an integral part of these research programs because it provides one of the most important means of observing global ecosystems consistently and synoptically.

Many scientific agencies throughout the world are focusing on monitoring the Earth to obtain accurate global climate change information. In this regard, NASA's Earth Science Program's goal is to develop a scientific understanding of Earth's system and its response to natural or human-induced changes, and to improve prediction of climate, weather, and natural hazards.

Early on, NASA conceptualized the remote sensing science conducted as part of its earth science program as consisting of two Earth subsystems: 1) the physical climate, and 2) biogeochemical cycles, linked by the global hydrologic cycle as shown in Figure 1-14 (Asrar and Dozier, 1994).

The **physical climate subsystem** is sensitive to fluctuations in the Earth's radiation balance. Many scientists believe that human activities have caused changes to

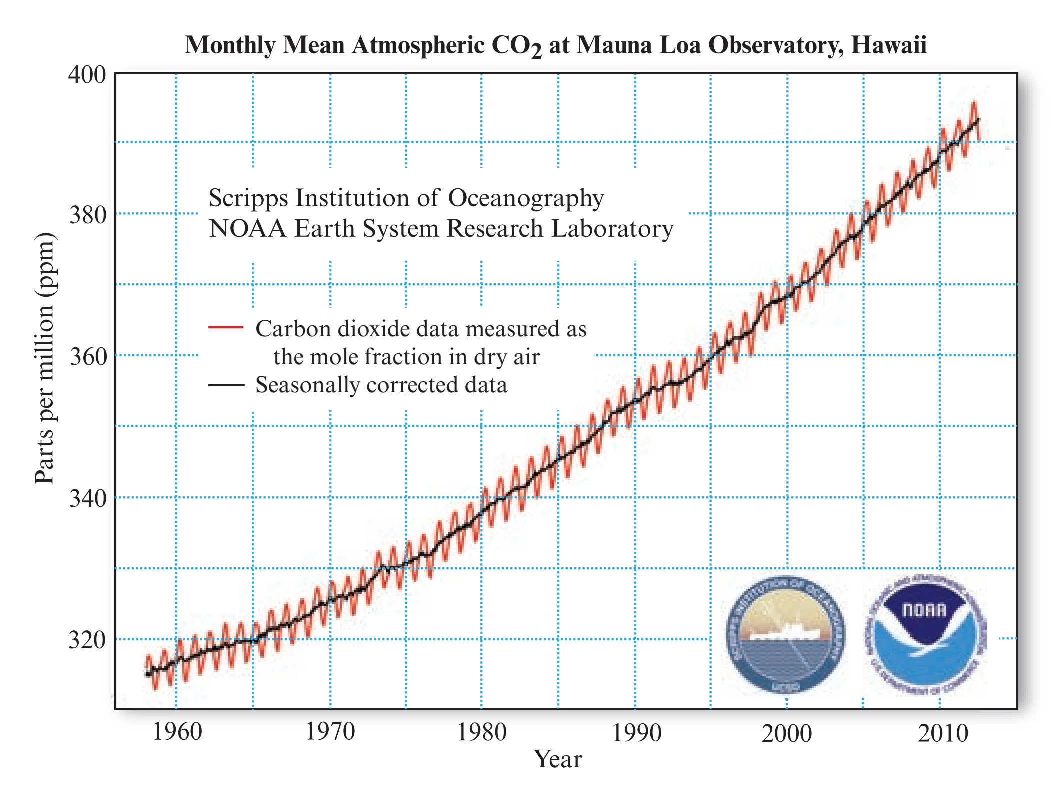


FIGURE 1–15 Monthly mean atmospheric carbon dioxide measured at Mauna Loa Observatory, Hawaii, from 1958 through October, 2012. Data are reported as a dry mole fraction defined as the number of molecules of carbon dioxide divided by the number of molecules of dry air multiplied by one million (ppm). Source of data: Scripps Institution of Oceanography and NOAA.

the planet's radiative heating mechanism that rival or exceed natural change. Most believe this to be due to an increase in atmospheric carbon dioxide. Mean monthly measurements of carbon dioxide at the Mauna Loa Observatory, HI, since 1958 are shown in Figure 1-15 (NOAA CO_2, 2012). If this rate is sustained, it could result in global mean temperatures increasing about 0.2 to 0.5 °C per decade during the next century. Volcanic eruptions and the ocean's ability to absorb heat may impact the projections.The following questions are being addressed using remote sensing:

- How do clouds, water vapor, and various types of aerosols in the Earth's radiation and heat budgets change with increased atmospheric greenhouse-gas concentrations?
- How do the oceans interact with the atmosphere in the transport and uptake of heat?
- How do land-surface properties such as snow and ice cover, evapotranspiration, urban/suburban land use, and vegetation influence circulation?

The Earth's **biogeochemical cycles** have also been changed by humans (Figure 1-14). Atmospheric carbon dioxide has increased by 30% since 1859, methane by more than 100%, and ozone concentrations in the stratosphere have decreased, causing increased levels of ultraviolet radiation to reach the Earth's surface. The following questions are being addressed using remote sensing:

- What role do the oceanic and terrestrial components of the biosphere play in the changing global carbon budget?
- What are the effects on natural and managed ecosystems of increased carbon dioxide and acid deposition, shifting precipitation patterns, and changes in soil erosion, river chemistry, and atmospheric ozone concentrations?

The **hydrologic cycle** links the physical climate and biogeochemical cycles (Figure 1-14). The phase change of water between its gaseous, liquid, and solid states involves storage and release of latent heat, so it influences atmospheric circulation and globally redistributes both water and heat (Asrar and Dozier, 1994). The hydrologic cycle is the integrating process for the fluxes of water, energy, and chemical elements among components of the Earth system. Important questions being addressed include:

- How will atmospheric variability, human activities, and climate change affect patterns of humidity, precipitation, evapotranspiration, and soil moisture?
- How does soil moisture vary in time and space?

- Can we predict changes in the global hydrologic cycle using present and future observation systems and models?

To address these questions, NASA launched several very important satellites as part of its Earth Observation program in the late 1990s and the first decade of the twenty-first century. Many of the sensors continue to provide valuable Earth resource information. In particular, the EOS *Terra* satellite was launched on December 18, 1999. It contained five remote sensing instruments described in Chapter 2, including the:

- *Moderate Resolution Imaging Spectrometer* (MODIS)
- *Advanced Spaceborne Thermal Emission and Reflection Radiometer* (ASTER),
- *Multiangle Imaging SpectroRadiometer* (MISR),
- *Clouds and the Earth's Radiant Energy System* (CERES), and
- *Measurements of Pollution in the Troposphere* (MOPITT).

The EOS *Aqua* satellite was launched in May, 2002 with an additional MODIS sensor.

More recently, NASA's Earth Science Division is coordinating the development of a series of satellite and airborne missions for long-term global observations of the land surface, biosphere, solid Earth, atmosphere, and oceans. The report titled *Responding to the Challenge of Climate and Environmental Change: NASA's Plan for a Climate-Centric Architecture for Earth Observations and Applications from Space* describes NASA's plan (NASA Goals, 2010). In particular, NASA is completing the development and launch of a set of a) Foundational missions, b) new Decadal Survey requested missions, and c) Climate Continuity missions (NASA Earth Science Program, 2013).

The **Foundational missions** are those missions that were already in development at the time the National Research Council's decadal survey was published in 2007, including:

- *Aquarius*,
- *NPOESS Preparatory Project* (NPP),
- *Landsat Data Continuity Mission* (LDCM), and
- *Global Precipitation Measurement* (GPM).

The National Research Council completed its first decadal survey for Earth science titled *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond* in 2007 at the request of NASA, NOAA, and USGS (NRC, 2007b). The report recommended that: "The U.S. government, working in concert with the private sector, academia, the public, and its international partners, should renew its investment in Earth-observing systems and restore its leadership in Earth science and applications." The **Decadal Survey missions** guided by the 2007 decadal survey include:

- *Soil Moisture Active-Passive* (SMAP),
- *Ice, Cloud and land Elevation Satellite* (ICESat-II),
- *Hyperspectral Infrared Imager* (HyspIRI),
- *Active Sensing of CO2 Emissions Over Nights, Days, and Seasons* (ASCENDS),
- *Surface Water and Topography* (SWOT),
- *Geostationary Coastal and Air Pollution Events* (GEO-CAPE), and
- *Aerosol-Clouds-Ecosystems* (ACE).

The **Climate Continuity missions** include:

- *Orbiting Carbon Observatory-2* (OCO-2),
- *Stratospheric Aerosol and Gas Experiment – III* (SAGE III), *Gravity Recovery and Climate Experiment Follow-on* (GRACE-FO), and
- *Pre-Aerosol, Clouds, and Ocean Ecosystem* (PACE) (NASA Earth Science Program, 2013).

Several of these missions are discussed in Chapter 2.

Commercial Satellite Remote Sensing Data Providers: Commercial vendors continue to develop some of the most sophisticated remote sensing satellites for Earth resource applications. For example, GeoEye, Inc., launched GeoEye-1 on September 6, 2008, with a 0.41 × 0.41 m panchromatic band and four 1.65 × 1.65 m multispectral bands. DigitalGlobe, Inc., launched WorldView-2 on October 8, 2009, with a 0.46 × 0.46 m panchromatic band and four 1.85 × 1.85 m multispectral bands. Astrium, Inc., launched Pleiades-1 on December 16, 2011, with four multispectral bands at 2 × 2 m and a panchromatic band at 0.5 × 0.5 m. RapidEye, Inc., launched RapidEye on August 29, 2008, with five multispectral bands at 5 × 5 m spatial resolution. ImageSat International, Inc., launched EROS-B on April 25, 2006, with a 0.7 × 0.7 m panchromatic band. SPOT Image launched SPOT 6 on September 9, 2012, with a 1.5 × 1.5 m panchromatic band and four multispectral bands at 8 × 8 m. The characteristics of these commercial satellite remote sensing systems are described in Chapter 2.

Remote Sensing Data Analysis

Remote sensor data are analyzed using a variety of image processing techniques (Figure 1-5), including:

- analog (visual) image processing, and
- digital image processing.

Analog and digital analysis of remotely sensed data seeks to detect and identify important phenomena in the scene. Once identified, the phenomena are usually measured, and the information is used in solving problems (Jensen and Hodgson, 2005; Roy et al., 2014). Thus, both manual and digital analysis have the same general goals. However, the attainment of these goals may follow different paths.

Human beings are adept at visually interpreting images produced by certain types of remote sensing devices, especially cameras. One could ask, "Why try to mimic or improve on this capability?" First, there are certain thresholds beyond which the human interpreter cannot detect "just noticeable differences" in the imagery. For example, it is commonly known that an analyst can discriminate only about nine shades of gray when interpreting continuous-tone, black-and-white photography. If the data were originally recorded with 256 shades of gray, there might be more subtle information present in the image than the interpreter can extract visually. Furthermore, the interpreter brings to the task all the pressures of the day, making the interpretation subjective and generally unrepeatable. Conversely, the results obtained by computer are repeatable (even when wrong!). Also, when it comes to keeping track of a great amount of detailed quantitative information, such as the spectral characteristics of a vegetated field throughout a growing season for crop identification purposes, the computer is very adept at storing and manipulating such tedious information and possibly making a more definitive conclusion as to what crop is being grown. Furthermore, digital image processing may improve efficiency and reduce the cost of expensive human labor. This is not to say that digital image processing is superior to visual image analysis. Rather, there may be times when a digital approach is better suited to the problem at hand. Optimum results are often achieved using a synergistic combination of both visual and digital image processing.

Analog (Visual) Image Processing

Humans use the fundamental **elements of image interpretation** summarized in Figure 1-16, including grayscale tone, color, height (depth), size, shape, shadow, texture, site, association, and arrangement. Most of these elements of image interpretation were first introduced by Olson (1960). The human mind is amazingly good at recognizing and associating these complex elements in an image or photograph because we constantly process a) profile views of Earth features every day, and b) images seen in books, magazines, the television, and the Internet. Furthermore, we are adept at bringing to bear all the knowledge in our personal background and collateral information. We then converge all this evidence to identify phenomena in images and judge their significance. Precise measurement of objects (length, area, perimeter, volume, etc.) may be performed using photogrammetric techniques applied to either monoscopic (single-photo) or stereoscopic (overlapping) images. Numerous books have been written on how to perform visual image interpretation and photogrammetric measurement.

There is a resurgence in the art and science of visual image interpretation as the digital remote sensor systems provide increasingly higher spatial resolution imagery. Many people are displaying high spatial resolution GeoEye-1 and WorldView-2 imagery on the computer screen and then visually interpreting the data. The data are also often used as a base map in GIS projects (Jensen and Jensen, 2013).

Digital Image Processing

Scientists have made advances in digital image processing of remotely sensed data for scientific visualization and hypothesis testing. Many of the methods are summarized in Lillesand et al. (2008), Jensen et al. (2009), Warner et al. (2009a), Bossler et al. (2010), Prost (2013) and in this book. Digital image processing now makes use of many elements of image interpretation using the techniques summarized in Figure 1-16. Some of the major types of digital image processing include image preprocessing (radiometric and geometric correction), image enhancement, photogrammetric image processing of stereoscopic imagery, parametric and non-parametric information extraction, expert system (e.g., decision-tree) and neural network image analysis, hyperspectral data analysis, and change detection (Figure 1-5).

Radiometric Correction: Analog and digital remotely sensed imagery may contain noise or error that was introduced by the sensor system (e.g., electronic noise) or the environment (e.g., atmospheric scattering of light into the sensor's field of view). Advances have been made in our ability to remove these deleterious effects through simple image normalization techniques and more advanced absolute radiometric calibration of the data to scaled surface reflectance (for optical data) (e.g., He and Chen, 2014). Calibrated remote sensor data allow imagery and derivative products obtained on different dates to be compared (e.g., to measure the change in LAI between two dates) (Liang, 2004). Fundamental radiometric correction digital image processing principles are discussed in Chapter 6.

Geometric Correction: Most analog and digital remote sensor data are now processed so that individual picture elements are in their proper planimetric (x, y) positions in a standard map projection (e.g., Shepherd et al., 2014). This facilitates the use of the imagery and derivative products in GIS or spatial decision support systems.

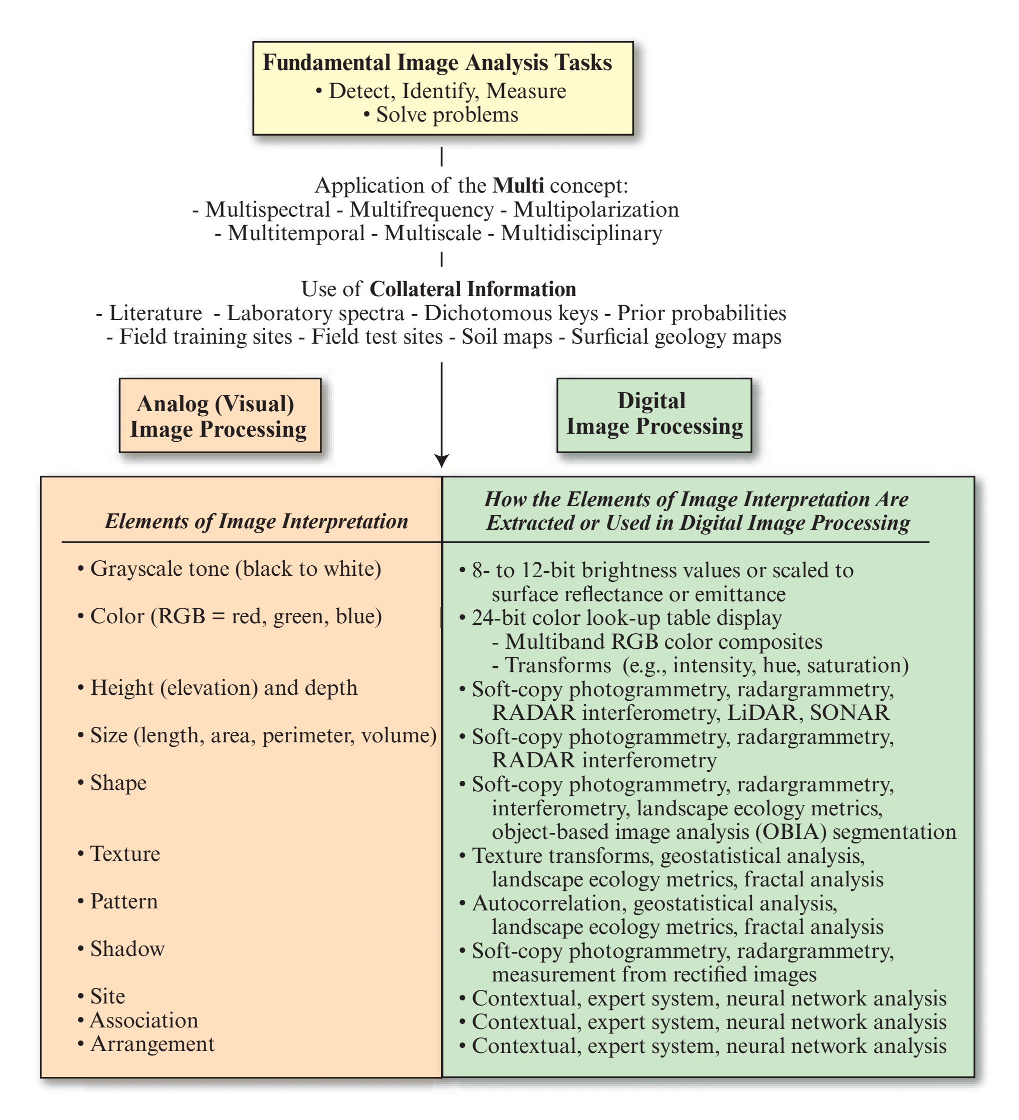


FIGURE 1–16 Analog (visual) and digital image processing of remotely sensed data make use of the fundamental elements of image interpretation.

Image Enhancement: Images can be digitally enhanced to identify subtle information in the analog or digital imagery that might otherwise be missed. Significant improvements have been made in our ability to contrast stretch and filter data to enhance low- and high-frequency components, edges, and texture in the imagery (e.g., Jensen and Jensen, 2013). In addition, the remote sensor data can be linearly and nonlinearly transformed into information that is more highly correlated with real-world phenomena through principal components analysis and various vegetation indices (Nellis et al., 2009).

Photogrammetry: Significant advances have been made in the analysis of stereoscopic remote sensor data obtained from airborne or satellite platforms using computer workstations and image processing photogrammetric algorithms. Soft-copy photogrammetric workstations can be used to extract accurate digital elevation models (DEMs) and differentially corrected orthophotography from the triangulated aerial photography or imagery (Leprince et al., 2007; Wolf et al., 2013). Large-scale mapping of building footprints, transportation network, hydrologic network, utilities, etc. is now performed almost exclusively using soft-copy photogrammetric and/or LiDARgrammetric data analysis (Jensen and Jensen, 2013).

Parametric Information Extraction: Scientists attempting to extract land-cover information from remotely sensed data now routinely specify if the classification is to be:

- **hard** (sometimes referred to as **crisp**), with discrete mutually exclusive classes, or **fuzzy**, where the proportions of materials within pixels are extracted;
- based on individual pixels (referred to as a **per-pixel classification**) or if it will use object-based image analysis (OBIA) segmentation algorithms that take into account not only the spectral characteristics of

a pixel, but also the spectral characteristics of contextual surrounding pixels. Thus, the algorithms take into account spectral and spatial information (e.g., Tullis and Jensen, 2003; Blaschke et al., 2014).

Once these issues are addressed, it is a matter of determining whether to use parametric (based on the analysis of normally distributed data), nonparametric, and/or nonmetric classification techniques. The maximum likelihood classification algorithm continues to be a widely used parametric classification algorithm. Unfortunately, the algorithm requires normally distributed training data in *n* bands (rarely the case) for computing the class variance and covariance matrices. It is difficult to incorporate nonimage categorical data into a maximum likelihood classification. Fortunately, fuzzy maximum likelihood classification algorithms are now available (e.g., Shackelford and Davis, 2003; Liu et al., 2011). Support Vector Machine (SVM) classification is also very effective especially when spectral training data consist of mixed pixels (Jensen et al., 2009).

Nonparametric Information Extraction: Nonparametric clustering algorithms, such as ISODATA, continue to be used extensively in digital image processing research. Unfortunately, such algorithms depend on how the seed training data are extracted and it is often difficult to label the clusters to turn them into useful information classes. For these and other reasons there has been a significant increase in the development and use of artificial neural networks (ANN) for remote sensing applications. An ANN does not require normally distributed training data. An ANN may incorporate virtually any type of spatially distributed data in the classification. The only drawback is that sometimes it is difficult to determine exactly how the ANN came up with a certain conclusion because the information is locked within the weights in the hidden layer(s). Scientists continue to work on ways to extract hidden information so that the rules used can be more formally stated. The ability of an ANN to learn should not be underestimated.

Nonmetric Information Extraction: It is difficult to make a computer understand and use the heuristic rules of thumb and knowledge that a human expert uses when interpreting an image. Nevertheless, there has been progress in the use of artificial intelligence (AI) to try to make computers do things that, at the moment, people do better. One area of AI that has great potential for image analysis is the use of expert systems that place all the information contained within an image in its proper context with ancillary data and extract valuable information. Duda et al. (2001) describe various types of expert system decision-tree classifiers as nonmetric.

Parametric digital image classification techniques are based primarily on summary statistics such as the mean, variance, and covariance matrices. Decision-tree or rule-based classifiers are not based on inferential statistics, but instead "let the data speak for itself" (Gahegan, 2003). In other words, the data retains its precision and is not dumbed down by summarizing it through means, etc. Decision-tree classifiers can process virtually any type of spatially distributed data and can incorporate prior probabilities (McIver and Friedl, 2002; Roberts et al., 2002; Im and Jensen, 2005). There are several approaches to rule creation, including: 1) explicitly extracting knowledge and creating rules from experts, 2) implicitly extracting variables and rules using cognitive methods (Lloyd et al., 2002), and 3) empirically generating rules from observed data and automatic induction methods (Tullis and Jensen, 2003). The development of a decision tree using human-specified rules is time-consuming and difficult. However, it rewards the user with detailed information about how individual classification decisions were made (Zhang and Wang, 2003).

Ideally, computers can derive the rules from training data without human intervention. This is referred to as machine-learning (Huang and Jensen, 1997). The analyst identifies representative training areas. The machine learns the patterns from these training data, creates the rules, and uses them to classify the remotely sensed data. The rules are available to document how decisions were made (Jensen et al., 2009). A drawback of artificial neural network and machine learning classifiers in general is the need for a large sample size of training data.

Hyperspectral: Special software is required to process hyperspectral data collected by imaging spectroradiometers such as AVIRIS and MODIS. Viper Tools (Roberts, 2014), MultiSpec (Landgrebe and Biehl, 2014), and the Environment for Visualizing Images - ENVI (Exelis ENVI, 2014) and others have pioneered the development of hyperspectral image analysis software. The software reduces the dimensionality of the data (number of bands) to a degree, while retaining the essence of the data. Under certain conditions the software can be used to compare the remotely sensed spectral reflectance curves with a library of spectral reflectance curves. Analysts are also able to identify the type and proportion of different materials within an individual picture element using end-member spectral mixture analysis (Pu et al., 2008; Roberts, 2014).

Modeling Using a GIS Approach: Remotely sensed data should not be analyzed in a vacuum without the benefit of collateral (ancillary) information such as soil maps, hydrology, and topography. For example, land-cover mapping using remotely sensed data has been

significantly improved by incorporating topographic information from digital terrain models and historical land-use information (e.g., Recio et al., 2011). GIS studies require timely, accurate updating of the spatially distributed variables in the database that remote sensing can provide (Jensen and Jensen, 2013). Remote sensing can benefit from access to accurate ancillary information to improve classification accuracy and other types of modeling (e.g., Coops et al., 2006; Cho, 2009; Pastick et al., 2011). Such synergy is critical if successful expert system and neural network analyses are to be performed.

Scene Modeling: Strahler et al. (1986) describe a framework for modeling in remote sensing. Basically, a remote sensing model has three components: 1) a scene model, which specifies the form and nature of the energy and matter within the scene and their spatial and temporal order; 2) an atmospheric model, which describes the interaction between the atmosphere and the energy entering and being emitted from the scene; and 3) a sensor model, which describes the behavior of the sensor in responding to the energy fluxes incident on it and in producing the measurements that constitute the image. They suggest that the problem of scene inference, then, becomes a problem of model inversion in which the order in the scene is reconstructed from the image and the remote sensing model. For example, Woodcock et al. (1997) inverted the Li-Strahler Canopy Reflectance Model for mapping forest structure. Deng et al. (2006) developed an algorithm to retrieve global leaf-area-index information.

Basically, successful remote sensing modeling predicts how much radiant flux in certain wavelengths should exit a particular object (e.g., a conifer canopy) even without actually sensing the object (Liang, 2009). When the model's prediction is the same as the sensor's measurement, the relationship has been modeled correctly. The scientist then has a greater appreciation for energy–matter interactions in the scene and may be able to extend the logic to other regions or applications with confidence. The remote sensor data can then be used more effectively in physical deterministic models (e.g., canopy structure, watershed runoff, net primary productivity, and evapotranspiration models) that are so important for large ecosystem modeling. (e.g., Disney et al., 2006).

Change Detection: Remotely sensed data obtained on multiple dates can be used to identify the type and spatial distribution of changes taking place in the landscape (Green, 2011). Digital change detection algorithms can provide binary urban land cover "change/no-change" information or more detailed "from-to" change information which identifies changes in land cover from forest, agriculture, etc., into urban land cover such as residential housing, apartment complexes or new roads (e.g., Jensen et al., 2009; Im et al., 2011; Tsai et al., 2011). The change information provides valuable insight into the processes at work (Jensen and Im, 2007; Purkis and Klemas, 2011). Change detection algorithms can be used on per-pixel and object-oriented (polygon) classifications. Unfortunately, there is still no universally accepted method of detecting change or of assessing the accuracy of change detection map products (Warner et al., 2009b).

Information Presentation

Information derived from remote sensor data are usually summarized as an enhanced image, image map, orthophotomap, thematic map, spatial database file, statistic, or graph (Figure 1-5). Thus, the final output products often require knowledge of remote sensing, cartography, GIS, and spatial statistics as well as the systematic science being investigated (e.g., soils, agriculture, urban studies). Scientists who understand the rules and synergistic relationships of the technologies can produce output products that communicate effectively. Those who violate fundamental rules (e.g., cartographic theory or database topology design) often produce poor output products that do not communicate effectively.

Image maps offer scientists an alternative to line maps for many cartographic applications. Thousands of satellite image maps have been produced from Landsat MSS (1:250,000 and 1:500,000 scale), TM (1:100,000 scale) and AVHRR, and MODIS data. Image maps at scales >1:24,000 are possible using imagery with a spatial resolution of < 1 × 1 m. Because image map products can be produced for a fraction of the cost of conventional line maps, they provide the basis for a national map series oriented toward the exploration and economic development of the less-developed areas of the world, most of which have not been mapped at scales of 1:100,000 or larger.

Remote sensor data that have been geometrically rectified to a standard map projection are becoming indispensable in most sophisticated GIS databases. This is especially true of orthophotomaps, which have the metric qualities of a line map and the information content of an aerial photograph or other type of image.

Unfortunately, when error is introduced into the remote sensing process it must be identified and reported. Innovations in error reduction include: 1) recording the lineage of the operations applied to the original remote sensor data, 2) documenting the geometric (spatial) error and thematic (attribute) error of the source materials, 3) improving legend design, especially for change detection map products derived from remote

FIGURE 1–17 Remote sensing Earth observation economics. The goal is to minimize the knowledge gap between the information delivery system, remote sensing experts, and the information consumer (user). The remote sensing–derived economic, social, strategic, environmental, and/or political information must be cost-effective, and easy to use to achieve equilibrium (adapted from Miller et al., 2003).

sensing, and 4) improved accuracy assessment. The remote sensing and GIS community should incorporate technologies that track all error in final map and image products as part of the metadata. This would result in more accurate information being used in the decision-making process.



Earth Observation Economics

The National Research Council recognized that there is an economic system at play when remote sensor data are used for Earth resource management applications (Figure 1-17) (Miller et al., 2001). It consists of an information delivery system with three components: data collection, image processing, and information consumer (user).

The data collection system is composed of commercial vendors and public agencies that operate remote sensing systems. Private industry provides information at market value. Public agencies generally provide remote sensor data at the cost of filling a user request (COFUR). Remote sensing has been around since the 1960s. There is an increasing number of experts that can use analog and/or digital image processing techniques to extract accurate geospatial information from the imagery. Then, there is the information consumer (user) of the remote sensing–derived information. The user generally needs information of economic, social, strategic, environmental and/or political value (NRC, 1998).

In order for the revenues generated by the information delivery system to be sufficient to support the capital and operating costs of the system, there must be a balance (equilibrium) between the value of the information, as perceived by the user (consumer), and the revenue necessary to support the system (Miller et al., 2001; 2003). The equilibrium has been achieved for airborne photogrammetric and LiDAR mapping applications for several decades. Time will tell if the balance between perceived value and cost can be maintained in the spaceborne case. Mergers are occurring.

The equilibrium can also be impacted by remote sensing technology experts who do not have a good understanding of the user information requirements. In fact, some remote sensing experts are often baffled as to why the lay consumers don't embrace the remote sensing–derived information. What they fail to consider is that the consumers generally have no motivation to switch to remote sensing–derived information on economic, social, environmental, strategic, or political attributes simply because it is based on new technology. Furthermore, the consumers on the right side of Figure 1-17 often have little knowledge of remote sensing technology or of how it is used to derive information.

TABLE 1–2 Selected remote sensing–related occupations and wage and employment trends defined by the U.S. Department of Labor Employment and Training Administration (USDOL/ETA, 2013; O*Net Online, http://online.onetcenter.org/find/quick?s=remote sensing, photogrammetry, and geographic information systems).

Code	Occupation	Wages & Employment Trends
19-2099.01	**Remote Sensing Scientists and Technologists**	Median wages (2013): $44.82 hour; $93,230 annual Employment (2012): 30,000 employees Projected growth (2012–2022): Slower than average (3 to 7%) Projected job openings (2012–2022): 8,300 Top industries (2012): 1. Government 2. Educational Services
19-4099.03	**Remote Sensing Technicians**	Median wages (2013): $21.25 hour; $44,200 annual Employment (2012): 64,000 employees Projected growth (2012–2022): Average (8 to 14%) Projected job openings (2012–2022): 31,600 Top industries (2012): 1. Government 2. Professional, Scientific, and Technical Services
17-1021.00	**Cartographers and Photogrammetrists**	Median wages (2013): $28.29 hour; $58,840 annual Employment (2012): 12,000 employees Projected growth (2012–2022): Faster than average (15 to 21%) Projected job openings (2012–2022): 4,900 Top industries (2012): 1. Professional, Scientific, and Technical Services 2. Government
15-1199.04	**Geographic Information Scientists and Technologists**	Median wages (2013): $39.59 hour; $82,340 annual Employment (2012): 206,000 employees Projected growth (2012–2022): Slower than average (3 to 7%) Projected job openings (2012–2022): 40,200 Top industries (2012): 1. Government 2. Professional, Scientific, and Technical Services

Notes:
Projected Growth is the estimated change in total employment over the projection period (2012–2022).
Projected Job Openings represent openings due to growth and replacement.
Industries are broad groups of businesses or organizations with similar activities, products, or services. Occupations are considered part of an industry based on their employment. There are hundreds of specific occupations within the "Government" and within the "Professional, Scientific, and Technical Service" industries. Detailed occupation lists can be viewed at http://www.onetonline.org/.

Miller et al. (2001; 2003) suggest that this situation can create a knowledge gap between the remote sensing experts and the information consumers (i.e., the users) (Figure 1-17). Bridging the gap is mandatory if we are to use remote sensing to solve Earth resource management problems. It is unlikely that the user community can devote the time to learn the physics of remote sensing and methods of analog and/or digital image processing and GIS modeling necessary to produce useful information. Conversely, there is considerable interest on the technology side of the problem to build a communication bridge. Therefore, one way to decrease the size of the knowledge gap is for the remote sensing technologists to work more closely with the user communities to understand their requirements. This will lead to more useful remote sensing–derived information of value to the user communities.

Advances in remote sensing image delivery systems by commercial firms such as Google's *Google Earth* and Microsoft's *Bing Maps* are having a significant impact on the public's use and appreciation of remote sensor data.



Remote Sensing/Digital Image Processing Careers in the Public and Private Sectors

Before forging ahead into the next chapter, it is useful to take a brief look at why knowledge of remote sensing and digital image processing might be of value to you. Currently, there are many GIScience employment opportunities in the public or private sectors. The O*NET program is the nation's primary source of geospatial occupational information (USDOL/ETA, 2014). The O*NET database contains information on >900 standardized occupation-specific descriptors (Table 1-2). It is available to the public at no cost and it is continually updated by surveying a broad range of workers from each occupation. A search of the O*Net database using the keywords "remote sensing, photogrammetry, and geographic information systems" identified several remote sensing–related occupations and their wage and employment trends through 2022.

Just a few of the remote sensing, photogrammetry, and GIS-related occupations are listed in Table 1-2.

It is clear from Table 1-2 that a) in 2012 there were 30,000 remote sensing scientists and technologists and 64,000 remote sensing technicians, b) there is significant demand for more people in the field from 2012 to 2022 (>39,000), with a projected growth in remote sensing–related occupations from 3% to 14%, and c) the remote sensing scientist occupations pay twice as well as the remote sensing technician positions. There were more than 206,000 GIS scientists and technologists in 2012 with a projected demand of 40,200 from 2012 to 2022. The top industries associated with the remote sensing–related occupations are "Government" "Educational Services," or "Professional, Scientific, or Technical Services."

A recent report titled *Future U.S. Workforce for Geospatial Intelligence* (NRC, 2013) predicts that people trained in remote sensing, digital image processing, photogrammetry, and other GIScience-related fields will continue to be in demand in the National Geospatial-Intelligence Agency (NGA) and the military.

Remote Sensing/Digital Image Processing Careers in the Public Sector

Public remote sensing and digital image processing employment will provide you with many opportunities to use remote sensing science for the public good and to interact with citizens on a regular basis. Public employees trained in remote sensing and digital image processing help monitor and protect natural resources, plan our cities, and monitor and maintain our infrastructure (roads, waterways, etc.). National Homeland Security and defense-related agencies such as the NGA, will likely continue to hire a great number of GIScience professionals in coming years (NRC, 2013).

Many people trained in remote sensing and digital image processing are drawn to public sector employment because of the perceived stability of employment. Although specific duties will vary by job, important aspects of remote sensing–related jobs will continue to be: a) procuring remote sensor data, b) analyzing remote sensor data to extract meaningful information, c) collecting *in situ* data in support of remote sensing missions, d) processing the remote sensing–derived information with other geospatial information in a GIS to perform predictive modeling, and e) maintaining large geospatial databases.

Working for a public agency will require you to provide regular reports/presentations to governmental organizations, such as city/county councils, planning agencies, public interest groups, etc. Therefore, verbal, written, and graphic communication skills are indispensable, particularly when describing the results and significance of remote sensing–related studies.

Unless "classified" as secret by the government, most of the data and projects that you work on in the public sector are available to the general public via the Freedom of Information Act. Public employees (except those in sensitive defense or homeland security related occupations) are generally allowed to publish the results of their remote sensing–related studies in popular or peer-reviewed literature. GIScience professionals working in public colleges and universities are expected to publish their research in peer-reviewed journals.

Remote Sensing/Digital Image Processing Careers in the Private Sector

Working for a private remote sensing and/or photogrammetric engineering firm will most likely require you to a) have a thorough knowledge of remote sensing and digital image processing, b) prepare proposals, often in response to a government or commercially sponsored Request for Proposal (RFP), and c) work on very specific projects with relatively tight schedules. Many private-sector employees are able to propose very specific remote sensing–related projects that are of great interest to them. To some this is an important private-sector consideration. Once again, verbal, written, and graphic communication skills are important.

Remote sensing-related data analysis results created while working in the private sector may be proprietary to the company or firm. You may or may not be allowed to publish your results in popular or refereed journal literature. Proprietary methods, procedures, and patents are the life-blood of many commercial firms who must maintain a competitive intellectual and economic advantage to survive.

To help prepare you to perform well in either the public or private sectors, you should have a sound knowledge of remote sensing and digital image processing fundamental principles and be proficient using the most widely adopted digital image processing, softcopy photogrammetry, and GIS software.



Earth Resource Analysis Perspective

Remote sensing is used for numerous applications such as medical imaging (e.g., x-raying a broken arm), nondestructive evaluation of products on an assembly line, and analysis of Earth resources. This book focuses on the art and science of applying remote sensing for the extraction of useful Earth resource information (Figure 1-18). Earth resource information is defined as any

Organization of
Introductory Digital Image Processing

Chapter 1.
Remote Sensing & Digital Image Processing
• Statement of the problem
• Select appropriate logic
• Select appropriate model

Chapter 2.
Remote Sensing Data Collection
• Digitize analog imagery, and/or
• Collect passive and active remote sensor data

Chapter 3.
Digital Image Processing Hardware and Software
• Configure software and hardware

Chapter 4.
Image Quality Assessment and Statistical Evaluation
• Univariate and multivariate statistics
• Geostatistics

Chapter 5.
Display Alternatives and Scientific Visualization
• 8-bit B&W and color look-up tables
• 24-bit color look-up tables

Chapter 6.
Electromagnetic Radiation Principles and Radiometric Correction
• Correcting sensor-induced radiometric error
• Atmospheric correction methods

Chapter 7.
Geometric Correction
• Image-to-image registration
• Image-to-map rectification
• Mosaicking

Chapter 8.
Image Enhancement
• Transects, contrast, spatial filtering
• Edge detection, texture
• Transformations

Chapter 9.
Information Extraction using Pattern Recognition
• Classification scheme
• Feature selection
• Supervised/unsupervised classification
• Fuzzy classification
• Object-based classification

Chapter 10.
Information Extraction using Artificial Intelligence
• Expert systems
• Artificial neural networks

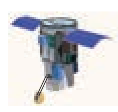
Chapter 11.
Information Extraction using Hyperspectral Image Analysis
• Radiometric/geometric correction
• Dimensionality reduction
• Endmember determination
• Mapping
• Matched filtering

Chapter 12.
Change Detection
• System and environmental variables
• Selection of appropriate algorithm

Chapter 13.
Accuracy Assessment
• Methods of accuracy assessment
• Sampling design
• Univariate/multivariate statistical analysis

FIGURE 1–18 This book is organized according to the remote sensing process (see Figure 1-5).

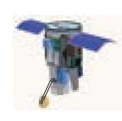
information concerning terrestrial vegetation, soils, minerals, rocks, water, and urban infrastructure as well as certain atmospheric characteristics. Such information may be useful for modeling the global carbon cycle, the biology and biochemistry of ecosystems, aspects of the global water and energy cycle, climate variability and prediction, atmospheric chemistry, characteristics of the solid Earth, population estimation, and monitoring land-use change and natural hazards (e.g., Mulder and Coops, 2014).



Book Organization

This book is organized according to the remote sensing process (Figures 1-5 and 1-18). The analyst first defines the problem and identifies the data required to accept or reject research hypotheses (Chapter 1). If a remote sensing approach to the problem is warranted, the analyst evaluates several data acquisition alternatives, including traditional aerial photography, multispectral scanners, and linear and area array multispectral and hyperspectral remote sensing systems (Chapter 2). For example, the analyst may digitize existing aerial photography or obtain the data already in a digital format from numerous public agencies (e.g., Landsat 8 data from the U. S. Geological Survey) or commercial firms (e.g., DigitalGlobe, Inc., RapidEye, Inc., Astrium Geoinformation Services, Inc., Itres, Inc., HyVista, Inc.). If the analysis is to be performed digitally, an appropriate digital image processing system is configured (Chapter 3). The image analysis begins by first computing fundamental univariate and multivariate statistics of the raw digital remote sensor data (Chapter 4). The imagery is then viewed on a computer screen or output to various hard-copy devices to analyze image quality (Chapter 5).

The imagery is then often preprocessed to reduce environmental and/or remote sensor system distortions. This preprocessing usually includes radiometric and geometric correction (Chapters 6 and 7). Various image enhancements may then be applied to the corrected data for improved visual analysis or as input to further digital image processing (Chapter 8). Thematic information may then be extracted from the imagery using pattern recognition, artificial intelligence, and/or hyperspectral image analysis techniques (Chapters 9–11). Multiple dates of imagery can be analyzed to identify change that provides insight into the processes at work (Chapter 12). Methods of assessing the accuracy of remote sensing–derived thematic map products are summarized in Chapter 13. Sources of remote sensing imagery and other selected geospatial information are summarized in the Appendix.



References

American Society for Photogrammetry & Remote Sensing, 1952, 1966, 2004, 2013 (6th Ed.), *Manual of Photogrammetry*, Bethesda: ASPRS.

Asrar, G., and J. Dozier, 1994, *EOS: Science Strategy for the Earth Observing System*, Woodbury, MA: American Institute of Physics, 342 p.

Azar, D., Engstrom, R., Graesser, J., and J. Comenetz, 2013, "Generation of Fine-scale Population Layers using Multi-Resolution Satellite Imagery and Geospatial Data," *Remote Sensing of Environment*, 130:219–232.

Blaschke, T., and 10 co-authors, 2014, "Geographic Object-Based Image Analysis — Towards A New Paradigm," *ISPRS Journal of Photogrammetry and Remote Sensing*, 87:180–191.

Bossler, J. D. (Ed.), 2010, *Manual of Geospatial Science & Technology*, 2nd Ed., London: Taylor & Francis, 832 p.

Brewin, R. J. W., Raitsos, D. E., Pradhan, Y., and I. Hoteit, 2013, "Comparison of Chlorophyll in the Red Sea Derived from MODIS-Aqua and *in vivo* Fluorescence," *Remote Sensing of Environment*, 136:218–224.

Cho, M., 2009, "Integrating Remote Sensing and Ancillary Data for Regional Ecosystem Assessment: *Eucalyptus grandis* Agro-system in KwaZulu-Natal, South Africa," *Proceedings of the IEEE Geoscience and Remote Sensing Symposium*, pages IV–264 to 267.

Colwell, R. N. (Ed.), 1960, *Manual of Photographic Interpretation*, Falls Church: ASPRS.

Colwell, R. N. (Ed.), 1983, *Manual of Remote Sensing*, 2nd Ed., Falls Church: ASPRS.

Colwell, R. N., 1997, "History and Place of Photographic Interpretation," in *Manual of Photographic Interpretation*, (2nd Ed.), W. Phillipson (Ed.), Bethesda: ASPRS, 33–48.

Congalton, R. G., and K. Green, 2009, *Assessing the Accuracy of Remotely Sensed Data—Principles and Practices* (2nd Ed.), Boca Raton: CRC Press, 183 p.

Coops, N. C., Wulder, M. A., and J. C. Whiete, 2006, *Integrating Remotely Sensed and Ancillary Data Sources to Characterize A Mountain Pine Beetle Infestation*, Victoria, CN: Natural Resources Canada, Canadian Forest Service, 33 p., http://cfs.nrcan.gc.ca/pubwarehouse/pdfs/26287.pdf.

Curran, P. J., 1987, "Remote Sensing Methodologies and Geography," *Intl. Journal of Remote Sensing*, 8:1255–1275.

Dahlberg, R. W., and J. R. Jensen, 1986, "Education for Cartography and Remote Sensing in the Service of an Information Society: The United States Case," *American Cartographer*, 13(1):51–71.

Deng, F., Chen, J. M., Plummer, S., Chen, M. Z., and J. Pisek, 2006, "Algorithm for Global Leaf Area Index Retrieval Us-

ing Satellite Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 44:2219–2229.

Disney, M., Lewis, P., and P. Saich, 2006, "3D Modeling of Forest Canopy Structure for Remote Sensing Simulations in the Optical and Microwave Domains," *Remote Sensing of Environment*, 100:114–132.

Dobson, J. E., 1993, "Commentary: A Conceptual Framework for Integrating Remote Sensing, Geographic Information Systems, and Geography," *Photogrammetric Engineering & Remote Sensing*, 59(10):1491–1496.

Duda, R. O., Hart, P. E., and D. G. Stork, 2001, *Pattern Classification*, NY: John Wiley, 394–452.

Estes, J. E. and J. R. Jensen, 1998, "Development of Remote Sensing Digital Image Processing Systems and Raster GIS," *History of Geographic Information Systems*, T. Foresman (Ed.), NY: Longman, 163–180.

Exelis ENVI, 2014, *Environment for Visualizing Images—ENVI Software*, Boulder: Exelis, Inc., http://www.exelisvis.com/ProductsServices/ENVIProducts/ENVI.aspx.

Fisher, P. F., and R. E. Lindenberg, 1989, "On Distinctions among Cartography, Remote Sensing, and Geographic Information Systems," *Photogrammetric Engineering & Remote Sensing*, 55(10):1431–1434.

Fussell, J., Rundquist, D., and J. A. Harrington, 1986, "On Defining Remote Sensing," *Photogrammetric Engineering & Remote Sensing*, 52(9):1507–1511.

Gahegan, M., 2003, "Is Inductive Machine Learning Just Another Wild Goose (or Might It Lay the Golden Egg)?" *Intl. Journal of GIScience*, 17(1):69–92.

Green, K., 2011, "Change Matters," *Photogrammetric Engineering & Remote Sensing*, 77(4):305–309.

He, Q., and C. Chen, 2014, "A New Approach for Atmospheric Correction of MODIS Imagery in Turbid Coastal Waters: A Case Study for the Pearl River Estuary," *Remote Sensing Letters,* 5(3):249–257.

Hodgson, M. E., Jensen, J. R., Raber, G., Tullis, J., Davis, B., Thompson, G., and K. Schuckman, 2005, "An Evaluation of LiDAR derived Elevation and Terrain Slope in Leaf-off Conditions," *Photogrammetric Engineering & Remote Sensing*, 71(7):817–823.

Huang, X., and J. R. Jensen, 1997, "A Machine Learning Approach to Automated Construction of Knowledge Bases for Image Analysis Expert Systems That Incorporate Geographic Information System Data," *Photogrammetric Engineering & Remote Sensing*, 63(10):1185–1194.

IGBP, 2014, *International Geosphere-Biosphere Programme*, Stockholm: IGBP at the Royal Swedish Academy of Sciences, http://www.igbp.net/.

Im, J., and J. R. Jensen, 2005, "A Change Detection Model based on Neighborhood Correlation Image Analysis and Decision Tree Classification," *Remote Sensing of Environment*, 99:326–340.

Im, J., Lu, Z., and J. R. Jensen, 2011, "A Genetic Algorithm Approach to Moving Threshold Optimization for Binary Change Detection," *Photogrammetric Engineering & Remote Sensing*, 77(2):167–180.

Irons, J. R., Dwyer, J. L., and J. A. Barsi, 2012, "The Next Landsat Satellite: The Landsat Data Continuity Mission," *Remote Sensing of the Environment*, 122:11–21.

Jensen, J. R., 1992, "Testimony on S. 2297, The Land Remote Sensing Policy Act of 1992," Senate Committee on Commerce, Science, and Transportation, Congressional Record, (May 6):55–69.

Jensen, J. R. and J. Im, 2007, "Remote Sensing Change Detection in Urban Environments," in R.R. Jensen, J. D. Gatrell and D. D. McLean (Eds.), *Geo-Spatial Technologies in Urban Environments Policy, Practice, and Pixels*, (2nd Ed.), Berlin: Springer-Verlag, 7–32.

Jensen, J. R., and M. E. Hodgson, 2005, "Chapter 8: Remote Sensing of Natural and Man-made Hazards and Disasters," in Ridd, M. K. and J. D. Hipple (Eds.), *Manual of Remote Sensing*: *Remote Sensing of Human Settlements,* (3rd Ed.), Bethesda: ASPRS, 401–429.

Jensen, J. R., Im, J. Hardin, P., and R. R. Jensen, 2009, "Chapter 19: Image Classification," in Warner, T. A., Nellis, M. D. and G. M. Foody (Eds.), *The Sage Handbook of Remote Sensing*, LA: Sage Publications, 269–281.

Jensen, J. R., and R. R. Jensen, 2013, *Introductory Geographic Information Systems*, Boston: Pearson, 400 p.

Jensen, J. R., and D. C. Cowen, 1999, "Remote Sensing of Urban/Suburban Infrastructure and Socioeconomic Attributes," *Photogrammetric Engineering & Remote Sensing*, 65(5):611–622.

Jensen, J. R., and R. E. Dahlberg, 1983, "Status and Content of Remote Sensing Education in the United States," *International Journal of Remote Sensing*, 4(2):235–245.

Jensen, J. R., and S. Schill, 2000, "Bi-directional Reflectance Distribution Function of Smooth Cordgrass (*Spartina alterniflora*)," *Geocarto International*, 15(2):21–28.

Jensen, J. R., Botchway, K., Brennan-Galvin, E., Johannsen, C., Juma, C., Mabogunje, A., Miller, R., Price, K., Reining, P., Skole, D., Stancioff, A., and D. R. Taylor, 2002, *Down to Earth: Geographic Information for Sustainable Development in Africa*, Washington: National Academy Press, 155 p.

Jensen, J. R., Qiu, F., and K. Patterson, 2001, "A Neural Network Image Interpretation System to Extract Rural and Urban Land Use and Land Cover Information from Remote Sensor Data," *Geocarto International*, 16(1):19–28.

Landgrebe, D., and L. Biehl, 2014, *MultiSpec*, W. Lafayette: Purdue University, https://engineering.purdue.edu/~biehl/MultiSpec/.

Leprince, S., Barbot, S., Ayoub, F., and J. Avouac, 2007, "Automatic and Precise Orthorectification, Coregistration, and Subpixel Correlation of Satellite Images, Application to

Ground Deformation Measurements," *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1529–1588.

Liang, S., 2004, *Quantitative Remote Sensing of Land Surfaces*, New York: John Wiley & Sons, 534 p.

Liang, S., 2009, "Quantitative Models and Inversion in Optical Remote Sensing," in Warner, T. A., Nellis, M. D. and G. M. Foody (Eds.), *The Sage Handbook of Remote Sensing*, LA: Sage Publications, 282–296.

Lillesand, T., Keifer, R., and J. Chipman, 2008, *Remote Sensing and Image Interpretation*, (6th Ed.), NY: John Wiley & Sons, 756 p.

Liu, K., Shi, W., and H. Zhang, 2011, "A Fuzzy Topology-based Maximum Likelihood Classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1):103–114.

Lloyd, R., Hodgson, M. E., and A. Stokes, 2002, "Visual Categorization with Aerial Photographs," *Annals of the Association of American Geographers.* 92(2):241–266.

Lu, D., and Q. Weng, 2004, "Spectral Mixture Analysis of the Urban Landscape in Indianapolis with Landsat ETM+ Imagery," *Photogrammetric Engineering & Remote Sensing*, 70(9):1053–1062.

McCoy, R., 2005, *Field Methods in Remote Sensing*, NY: Guilford, 159 p.

McIver, D. K., and M. A. Friedl, 2002, "Using Prior Probabilities in Decision-tree Classification of Remotely Sensed Data," *Remote Sensing of Environment*, 81:253–261.

Miller, R. B., Abbott, M. R., Harding, L. W., Jensen, J. R., Johannsen, C. J., Macauley, M., MacDonald, J. S., and J. S. Pearlman, 2001, *Transforming Remote Sensing Data into Information and Applications*, Washington: National Academy Press, 75 p.

Miller, R. B., Abbott, M. R., Harding, L. W., Jensen, J. R., Johannsen, C. J., Macauley, M., MacDonald, J. S., and J. S. Pearlman, 2003, *Using Remote Sensing in State and Local Government: Information for Management and Decision Making*, Washington: National Academy Press, 97 p.

NASA AVIRIS, 2014, *Airborne Visible/Infrared Imaging Spectrometer–AVIRIS*, Pasadena: NASA Jet Propulsion Laboratory, http://aviris.jpl.nasa.gov/.

NASA Earth Science Program, 2013, *NASA Earth Science Program*, Washington: NASA, http://nasascience.nasa.gov/earth-science/.

NASA Goals, 2010, *Responding to the Challenge of Climate and Environmental Change: NASA's Plan for a Climate-Centric ARchitecture for Earth Observations and Applications from Space*, Washington: NASA, 49 p.

Nellis, M. D., Price, K. P., and D. Rundquist, 2009, "Remote Sensing of Cropland Agriculture," Chapter 26 in Warner, T. A., Nellis, M. D. and G. M. Foody (Eds.), 2009, *The Sage Handbook of Remote Sensing*, Los Angeles: Sage Publications, 368–380.

NOAA CO2, 2012, *Trends in Atmospheric Carbon Dioxide*, Washington: NOAA Earth System Research Laboratory, http://www.esrl.noaa.gov/gmd/ccgg/trends/co2_data_mlo.html.

NRC, 1998, *People and Pixels: Linking People and Social Science*, Washington: National Academy Press, 244 p.

NRC, 2007a, *Contributions of Land Remote Sensing for Decisions About Food Security and Human Health: Workshop Report*, Washington: National Academy Press, 230 p.

NRC, 2007b, *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*, Washington: National Academy Press, 72 p.

NRC, 2009, *Uncertainty Management in Remote Sensing of Climate Date*, Washington: National Academy Press, 64 p.

NRC, 2012, *Ecosystem Services: Charting A Path to Sustainability*, Washington: National Academy Press, 136 p.

NRC, 2013, *Future U.S. Workforce for Geospatial Intelligence*, Washington: National Academy Press, 169 p.

NRC, 2014, *Report in Brief: Responding to Oil Spills in the U.S. Arctic Marine Environment*, Washington: National Academy Press, 7 p.

Olson, C. E., 1960, "Elements of Photographic Interpretation Common to Several Sensors," *Photogammetric Engineering*, 26(40):651–656.

Pastick, N., Wylie, B. K., Minsley, B. J., Jorgensor, T., Ji, L., Walvoord, A., Smith, B. D., Abraham, J. D., and J. Rose, 2011, "Using Remote Sensing and Ancillary Data to Extend airborne Electromagnetic Resistivity Surveys for Regional Permafrost Interpretation," paper presented at the *American Geophysical Union Fall Technical Meeting*, http://adsabs.harvard.edu/abs/2011 AGUFM.C41B0390P.

Prost, G. L., 2013, *Remote Sensing for Geoscientists—Image Analysis and Integration,* (3rd Ed.), Boca Raton: CRC Press, 702 p.

Pruitt, E. L., 1979, "The Office of Naval Research and Geography," *Annals*, Association of American Geographers, 69(1):106.

Pu, R., Gong, P., Michishita, R., and T. Sasagawa, 2008, "Spectral Mixture Analysis for Mapping Abundance of Urban Surface Components from the Terra/ASTER Data," *Remote Sensing of Environment*, 112:939–954.

Purkis, S., and V. Klemas, 2011, *Remote Sensing and Global Environmental Change*, NY: Wiley-Blackwell, 367 p.

Qiu, F., and J. R. Jensen, 2005, "Opening the Neural Network Black Box and Breaking the Knowledge Acquisition Bottleneck of Fuzzy Systems for Remote Sensing Image Classification," *International Journal of Remote Sensing*, 25(9):1749–1768.

Recio, J. A., Hermosilla, L., Ruiz, A., and A. Hernandez-Sarria, 2011, "Historical Land Use as a Feature for Image Classification," *Photogrammetric Engineering & Remote Sensing*, 77(4):377–387.

Renslow, M. (Ed.), 2012, *ASPRS Airborne Topographic LiDAR Manual*, Bethesda: ASPRS, 528 p.

Roberts, D. A., 2014, V*IPER Tools Version 1.5,* Santa Barbara: Univ of California Santa Barbara, Department of Geography, http://www.vipertools.org/.

Roberts, D. A., Numata, I., Holmes, K., Batista, G., Krug, T., Monteiro, A., Powell, B., and O. A. Chadwick, 2002, "Large Area Mapping of Land-cover Change in Rondônia using Multitemporal Spectral Mixture Analysis and Decision Tree Classifiers," *J. Geophys. Res.*, 107(D20), 8073, doi:10.1029/2001JD000374.

Roy, D. P., and 33 co-authors, 2014, "Landsat-8: Science and Product Vision for Terrestrial Global Change Research," *Remote Sensing of Environment*, 145:154–172.

Sandmeier, S. R., 2000, "Acquisition of Bidirectional Reflectance Factor Data with Field Goniometers," *Remote Sensing of Environment*, 73:257–269.

Schill, S., Jensen, J. R., Raber, G., and D. E. Porter, 2004, "Temporal Modeling of Bidirectional Reflection Distribution Function (BRDF) in Coastal Vegetation," *GIScience & Remote Sensing*, 41(2):116–135.

Shackelford, A. K., and C. H. Davis, 2003, "A Hierarchial Fuzzy Classification Approach for High-Resolution Multispectral Data Over Urban Areas," *IEEE Transactions on Geoscience and Remote Sensing,* 41(9):1920–1932.

Shepherd, J. D., Dymond, J. R., Gillingham, S., and P. Bunting, 2014, "Accurate Registration of Optical Satellite Imagery with Elevation Models for Topographic Correction," *Remote Sensing Letters*, 5(7):637–641.

Skidmore, A. K., 2002, "Chapter 2: Taxonomy of Environmental Models in the Spatial Sciences," in *Environmental Modelling with GIS and Remote Sensing*, A. K. Skidmore (Ed.), London: Taylor & Francis, 8–25.

Strahler, A. H., Woodcock, C. E., and J. A. Smith, 1986, "On the Nature of Models in Remote Sensing," *Remote Sensing of Environment*, 20:121–139.

Tsai, Y. H., Stow, D., and J. Weeks, 2011, "Comparison of Object-Based Image Analysis Approaches to Mapping New Buildings in Accra, Ghana Using Multi-Temporal QuickBird Satellite Imagery," *Remote Sensing*, 2011(3):2707–2726.

Tullis, J. A., and J. R. Jensen, 2003, "Expert System House Detection in High Spatial Resolution Imagery Using Size, Shape, and Context," Geocarto International, 18(1):5–15.

USDOL/ETA, 2014, *O*Net Online*, Washington: U.S. Department of Labor/Employment and Training Administration (www.onetonline.org/find/quick?s= remote sensing).

USGCRP, 2014, *United States Global Change Research Program*, Washington, DC: USGCRP, http://www.globalchange.gov/.

Warner, T. A., Almutairi, A., and J. Y. Lee, 2009a, "Remote Sensing of Land Cover Change," in Warner, T. A., Nellis, M. D. and G. M. Foody (Eds.), *The Sage Handbook of Remote Sensing*, LA: Sage Publications, 459–472.

Warner, T. A., Nellis, M. D., and G. M. Foody (Eds.), 2009b, *The Sage Handbook of Remote Sensing*, LA: Sage Publications, 532 p.

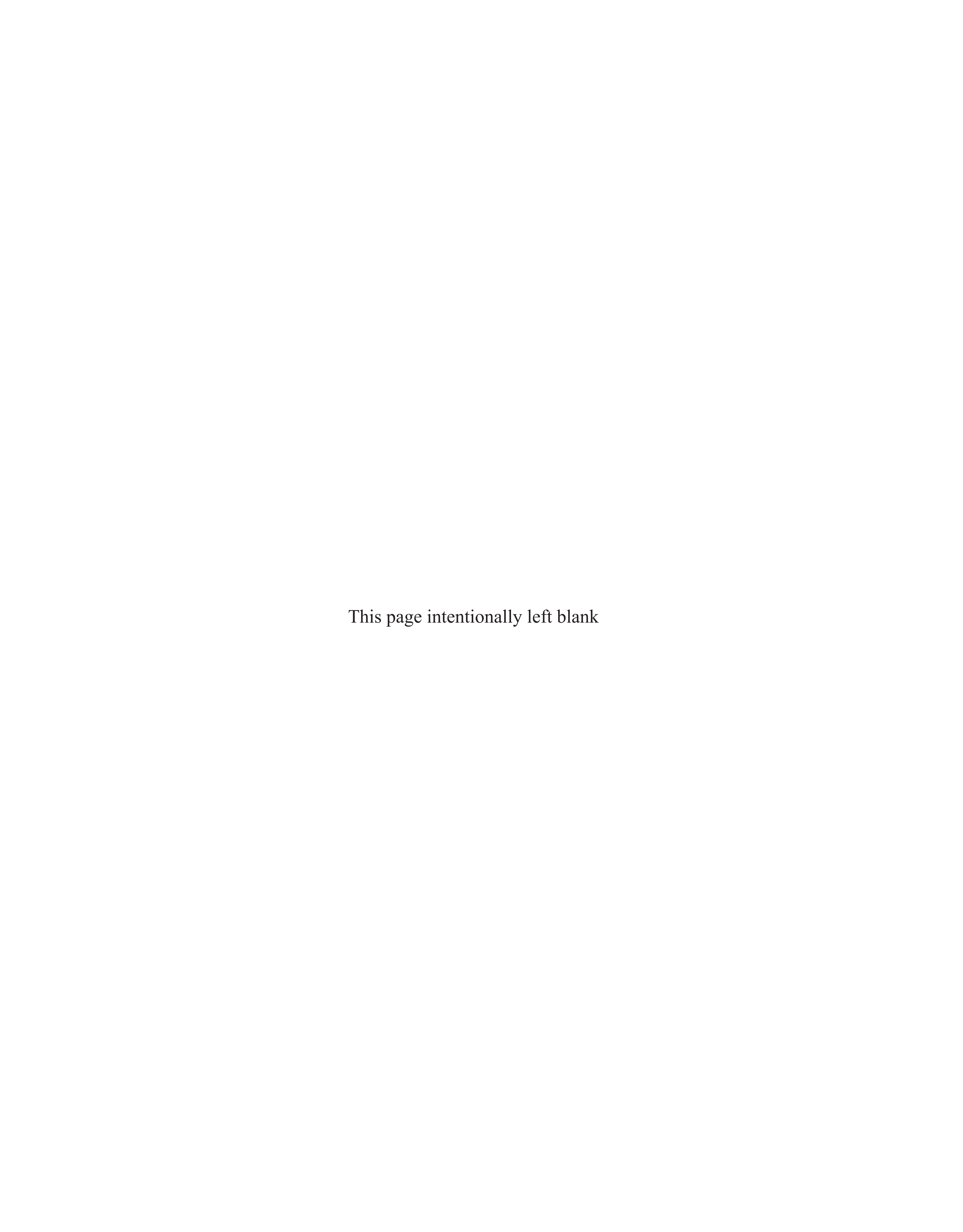
Wolf, P. R., Dewitt, R. A., and B. E. Wilkinson, 2013, *Elements of Photogrammetry with Applications in GIS*, 4th Ed., New York: McGraw–Hill.

Wolter, J. A., 1975, *The Emerging Discipline of Cartography*, Minneapolis: University of Minnesota, Department of Geography, unpublished dissertation.

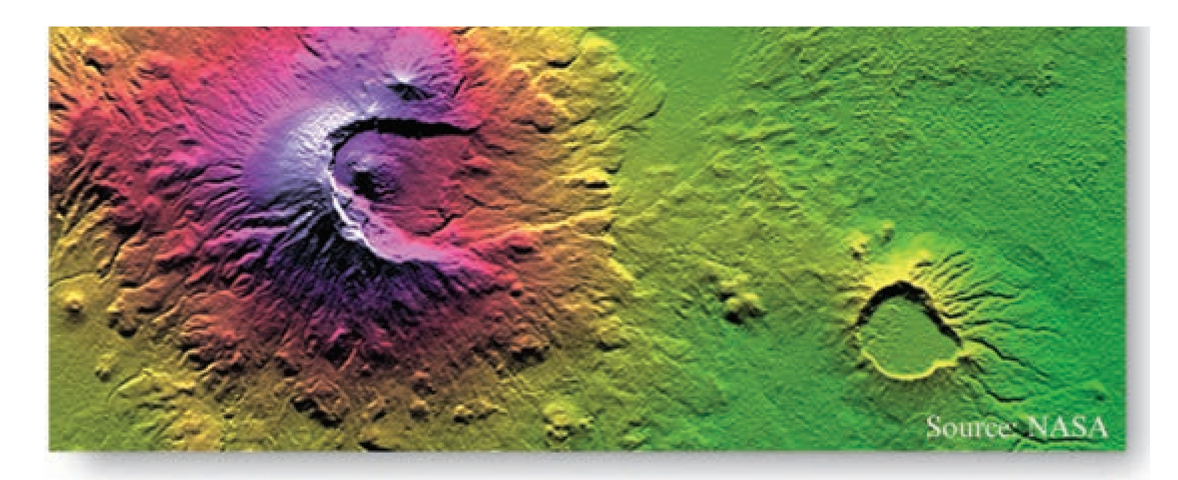
Woodcock, C. E., Collins, J. B., Jakabhazy, V., Li, X., Macomber, S., and Y. Wu, 1997, "Inversion of the Li-Strahler Canopy Reflectance Model for Mapping Forest Structure," *IEEE Transactions Geoscience & Remote Sensing*, 35(2):405–414.

Wulder, M. A., and N. C. Coops, 2014, "Make Earth Observations Open Access," *Nature*, 513:30-31.

Zhang, Q., and J. Wang, 2003, "A Rule-based Urban Land Use Inferring Method for Fine-resolution Multispectral Imagery," *Canadian Journal of Remote Sensing*, 29(1):1–13.


This page intentionally left blank

2 Remote Sensing Data Collection



Remotely sensed data must be in a digital format to perform digital image processing. There are two fundamental ways to obtain digital imagery:

1. acquire remotely sensed imagery in an *analog* format (often referred to as hard-copy) and then convert it to a digital format through the process of digitization, or
2. acquire remotely sensed imagery already in a *digital* format, such as that obtained by the Landsat 7 Enhanced Thematic Mapper Plus (ETM^+) sensor system, Landsat 8, Pleiades, or WorldView-3.



Overview

The chapter first introduces digital image terminology and how analog imagery is converted into digital imagery using digitization. It then provides detailed information about public and commercial satellite remote sensing systems (e.g., Landsat 7 and 8, SPOT 5–7, NPOESS Preparatory Project, Pleiades, GeoEye-1, WorldView-2 and -3, RADARSAT-1 and suborbital remote sensing systems that are often used to collect panchromatic, multispectral and hyperspectral imagery (e.g., Pictometry, AVIRIS, CASI 1500). The chapter provides details about the sensor system hardware, including: scanning systems, linear arrays, area arrays, and digital frame cameras. Formats for storing digital remote sensor data are described, including: band sequential, band interleaved by line, and band interleaved by pixel.



Analog (Hard-Copy) Image Digitization

Scientists and laypersons often obtain analog (hard-copy) remote sensor data that they desire to analyze using digital image processing techniques. Analog aerial photographs are, of course, ubiquitous because much of the world has been photographed many times. Occasionally scientists encounter hard-copy thermal-infrared or active microwave (RADAR) imagery. To convert analog imagery into digital imagery, the person performing the digital image processing must first understand digital image terminology.

Digital Image Terminology

Digital remote sensor data are usually stored as a matrix (array) of numbers. Each digital value is located at a specific row (*i*) and column (*j*) in the matrix (Figure 2-1). A **pixel** is defined as "a two-dimensional picture element that is the smallest non-divisible element of a digital image." Each pixel at row (*i*) and column (*j*) in the image has an original **brightness value** (*BV*) associated with it (some scientists use the term *digital number* [*DN*] value). The dataset may consist of *n* individual bands (*k*) of multispectral or hyperspectral imagery. Thus, it is possible to identify the brightness value of a particular pixel in the dataset by specifying its row (*i*), column (*j*), and band (*k*) coordinate, i.e., $BV_{i,j,k}$. It is important to understand that the *n* bands are all geometrically registered to one another. Therefore, a road intersection in band 1 at row 3, column 3 (i.e., $BV_{3,3,1}$) should be located at the same row and column coordinate in the fourth band (i.e., $BV_{3,3,4}$).

Digital Image Terminology and Radiometric Resolution

Picture element (pixel) at location row 3, column 3, band 1 has a brightness value of 20, i.e., $BV_{3,3,1} = 20$.

FIGURE 2–1 Digital remote sensor data are stored in a matrix (raster) format. Picture element (pixel) brightness values *(BV)* are located at row *i*, column *j*, and band *k* in the multispectral or hyperspectral dataset. The digital remote sensor brightness values are often stored as 8-bit bytes with values ranging from 0 to 255. However, several image digitization systems and some remote sensing systems now routinely collect 10-, 11-, or 12-bit data.

In an analog image, the brightness value (BV) is a surrogate for the *density* (D) of the light-absorbing silver or dye deposited at a specific location. The density of a negative or positive transparency film is measured using a **densitometer**. There are several types of densitometers, including flatbed and drum microdensitometers, video densitometers, and linear or area array charge-coupled-device densitometers.

Microdensitometer Digitization

The characteristics of a typical *flatbed microdensitometer* are shown in Figure 2-2. This instrument can measure the density characteristics of very small portions of a negative or positive transparency, down to just a few micrometers in size, hence the term *microdensitometer*. Basically, a known quantity of light is sent from the light source toward the receiver. If the light encounters a very dense portion of the film, very little light is transmitted to the receiver. If the light encounters a very clear portion of the film, then much of the light is transmitted to the receiver. The densitometer can output the characteristics at each i,j location in the photograph in terms of transmittance, opacity, or density. The ability of a portion of a *developed* film to pass light is called its *transmittance* ($\tau_{i,j}$). A black portion of the film may transmit no light, while a clear portion of the film may transmit almost 100 percent of the incident light. Therefore, the transmittance at location i,j in the photograph is:

$$\tau_{i,j} = \frac{\text{light passing through the film}}{\text{total incident light}}. \tag{2.1}$$

There is an inverse relationship between transmittance and the opacity of an area on the film. An area in the film that is very opaque does not transmit light well. *Opacity* ($O_{i,j}$) is the reciprocal of transmittance:

$$O_{i,j} = \frac{1}{\tau_{i,j}}. \tag{2.2}$$

Transmittance and opacity are two good measures of the darkness of any portion of a developed negative. However, psychologists have found that the human visual system does not respond linearly to light stimulation, but rather we respond logarithmically. Therefore, it is common to use *density* ($D_{i,j}$), which is the common logarithm of opacity, as the digitization measure of choice:

$$D_{i,j} = \log_{10} O_{i,j} = \log\left(\frac{1}{\tau_{i,j}}\right). \tag{2.3}$$

If 10% of the light can be transmitted through a film at a certain i,j location, transmittance is 1/10, opacity is 1/0.10 or 10, and density is the common logarithm of 10 or 1.0.

As previously mentioned, the amount of light recorded by the receiver in densitometric units is commonly converted into a digital brightness value ($BV_{i,j,k}$), which refers to the location in the photograph at row *i*, column *j*, and band *k*. At the end of each scan line, the light source steps in the *y*-direction some Δy to scan along a line contiguous and parallel to the previous one. As the light source is scanned across the image,

FIGURE 2–2 Schematic of a flatbed microdensitometer. The hard-copy remotely sensed imagery (usually a positive transparency) is placed on the flatbed surface. A small light source (perhaps 10 μm in diameter) is moved mechanically across the flat imagery in the *x*-direction, emitting a constant amount of light. On the other side of the imagery, a receiver measures the amount of energy that passes through. When one line scan is complete, the light source and receiver step in the *y*-direction some Δy to scan an area contiguous and parallel to the previous scan line. The amount of energy detected by the receiver along each scan line is eventually changed from an electrical signal into a digital value through an analog-to-digital (A-to-D) conversion. After the entire image has been scanned in this fashion, a matrix (raster) of brightness values is available for digital image processing purposes. A color filter wheel may be used if the imagery has multiple dye layers that must be digitized. In this case the imagery is scanned three separate times using three different filters to separate it into its respective blue, green, and red components. The resulting three matrices should be in near-perfect registration, representing a multispectral digital dataset (from Jensen and Jensen, 2013).

the continuous output from the receiver is converted to a series of discrete numerical values on a pixel-by-pixel basis. This analog-to-digital (A-to-D) conversion process results in a matrix of values that is usually recorded in 8-bit bytes (values ranging from 0 to 255). These data are then stored on disk or tape for future analysis.

Table 2-1 summarizes the relationship between digitizer scanning spot size (IFOV) measured in dots-per-inch (DPI) or micrometers (μm) and the pixel ground resolution at various scales of aerial photography or imagery. The algorithms for converting from DPI to μm and vice versa are also provided. Please remember that scanning imagery at spot sizes <12 μm may result in noisy digitized data because the spot size approaches the dimension of the film's silver halide crystals.

A simple black-and-white photograph has only a single band, $k = 1$. However, we may need to digitize color photography. In such circumstances, we use three specially designed filters that determine the amount of light transmitted by each of the dye layers in the film (Figure 2-2). The negative or positive color transparency is scanned three times (k = *1, 2,* and *3*), each time with a different filter. This extracts spectral information from the respective dye layers found in color and color-infrared aerial photography and results in a co-registered three-band digital dataset for subsequent image processing.

Rotating-drum optical-mechanical scanners digitize the imagery in a different fashion (Figure 2-3). The film transparency is mounted on a glass rotating drum so that it forms a portion of the drum's circumference. The light source is situated in the interior of the drum. The *y*-coordinate scanning motion is provided by the rotation of the drum. The *x*-coordinate is obtained by the incremental translation of the source-receiver optics after each drum revolution.

Flatbed and rotating drum microdensitometers yield the most accurate raster digitization of maps and images. Microdensitometers are often found at labs that conduct research to very exacting standards (e.g., softcopy photogrammetry) or at service companies that specialize in providing customers with high-quality raster digitization.

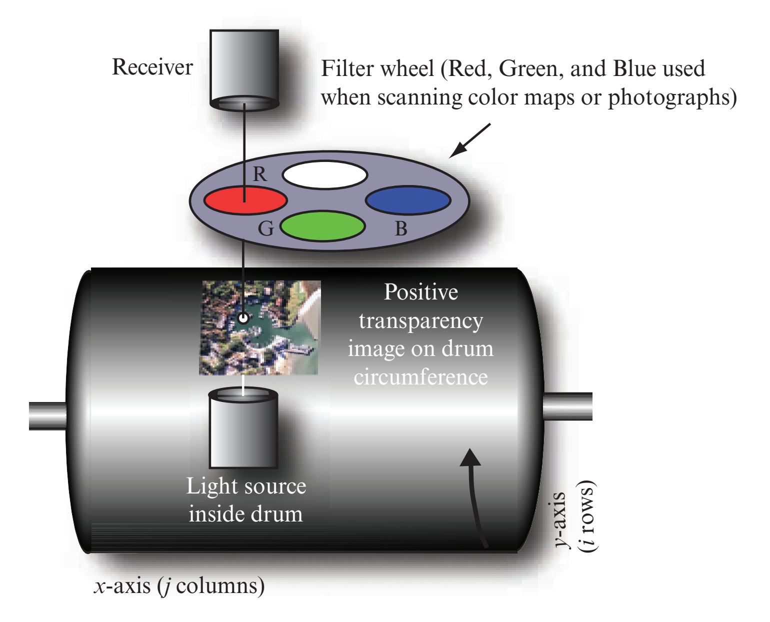


FIGURE 2–3 The rotating-drum, optical-mechanical scanner works on exactly the same principle as the flatbed microdensitometer except that the remotely sensed data are mounted on a rotating drum so that they form a portion of the drum's circumference. The light source is situated in the interior of the drum, and the drum is continually rotated in the *y*-direction. The *x*-coordinate is obtained by the incremental translation of the source-receiver optics after each drum revolution. Some microdensitometers can write to film as well as digitize from film. In such cases the light source (usually a photodiode or laser) is modulated such that it exposes each picture element according to its brightness value. These are called film-writers and provide excellent hard-copy of remotely sensed data (from Jensen and Jensen, 2013).

Video Digitization

It is possible to digitize hard-copy imagery by sensing it through a video camera and then performing an analog-to-digital conversion on the 525 columns by 512 rows of data that are within the standard field-of-view (as established by the National Television System Committee). Video digitizing involves freezing and then digitizing a frame of analog video camera input. A full frame of video input can be read in approximately 1/60 sec. A high-speed analog-to-digital converter, known as a *frame grabber*, digitizes the data and stores it in frame buffer memory. The memory is then read by the host computer and the digital information is stored on disk or tape.

Video digitization of hard-copy imagery is performed very rapidly, but the results are not always useful for digital image processing purposes due to:

- differences in the radiometric sensitivity of various video cameras, and
- vignetting (light fall-off) away from the center of the image being digitized.

These characteristics can affect the spectral data extracted from the imagery. Also, any geometric distortion in the vidicon optical system will be transferred to the digital remote sensor data, making it difficult to edge-match between adjacent images digitized in this manner.

Linear and Area Array Charge-Coupled-Device Digitization

Advances in the personal computer industry have spurred the development of flatbed, desktop linear array digitizers based on linear array charge-coupled devices (CCDs) that can be used to digitize hard-copy negatives, paper prints, or transparencies at 50 to 6,000 pixels per inch (Figure 2-4a–c). The hard-copy photograph is placed on the glass. The digitizer optical system illuminates an entire line of the hard-copy photograph at one time with a known amount of light. A linear array of detectors records the amount of light reflected from or transmitted through the photograph along the array and performs an analog-to-digital conversion. The linear array is stepped in the *y*-direction, and another line of data is digitized.

It is possible to purchase useful desktop color scanners for less than $200. Many digital image processing laboratories use these inexpensive desktop digitizers to convert hard-copy remotely sensed data and maps into a digital format. Desktop scanners provide surprisingly good spatial precision. An optional "transilluminator" can be purchased for back-lighting any transparency that needs to be scanned. Unfortunately, most desktop

TABLE 2–1 Relationship between digitizer detector instantaneous-field-of-view (IFOV) measured in dots-per-inch (DPI) or micrometers (μm), and the pixel ground resolution at various scales.

Digitizer Detector IFOV		Pixel Ground Resolution at Various Scales (in meters)					
Dots-per-inch (DPI)	**Micrometers (μm)**	**1:40,000**	**1:20,000**	**1:9,600**	**1:4,800**	**1:2,400**	**1:1,200**
100	**254.00**	10.16	5.08	2.44	1.22	0.61	0.30
200	**127.00**	5.08	2.54	1.22	0.61	0.30	0.15
300	**84.67**	3.39	1.69	0.81	0.41	0.20	0.10
400	**63.50**	2.54	1.27	0.61	0.30	0.15	0.08
500	**50.80**	2.03	1.02	0.49	0.24	0.12	0.06
600	**42.34**	1.69	0.85	0.41	0.20	0.10	0.05
700	**36.29**	1.45	0.73	0.35	0.17	0.09	0.04
800	**31.75**	1.27	0.64	0.30	0.15	0.08	0.04
900	**28.23**	1.13	0.56	0.27	0.14	0.07	0.03
1000	**25.40**	1.02	0.51	0.24	0.12	0.06	0.03
1200	**21.17**	0.85	0.42	0.20	0.10	0.05	0.03
1500	**16.94**	0.67	0.34	0.16	0.08	0.04	0.02
2000	**12.70**	0.51	0.25	0.12	0.06	0.03	0.02
3000	**8.47**	0.33	0.17	0.08	0.04	0.02	0.01
4000	**6.35**	0.25	0.13	0.06	0.03	0.02	0.008

Useful Scanning Conversions:

DPI = dots per inch; μm = micrometers; I = inches; M = meters
From DPI to micrometers: μm = (2.54 / DPI) 10,000
From micrometers to DPI: DPI = (2.54 / μm) 10,000
From inches to meters: M = I × 0.0254
From meters to inches: I = M × 39.37

Computation of Pixel Ground Resolution:

PM = pixel size in meters; PF = pixel size in feet; S = photo or map scale factor
Using DPI: PM = (S/DPI)/39.37 PF = (S/DPI)/12
Using micrometers: PM = (S × μm) 0.000001PF = (S × μm) 0.00000328
For example, if a 1:6,000 scale aerial photograph is scanned at 500 DPI, the pixel size will be (6000/500)/39.37 = 0.3048 meters per pixel or (6000/500)/12 = 1.00 foot per pixel. If a 1:9,600 scale aerial photograph is scanned at 50.8 μm, the pixel size will be (9,600 × 50.8)(0.000001) = 0.49 meters or (9,600 × 50.8)(0.00000328) = 1.6 feet per pixel.

scanners are designed for 8.5 × 14 in. originals, and most aerial photographs are 9 × 9 in. Similarly, most hard-copy maps are greater than 8.5 × 14 in. Under such conditions, the analyst must digitize the 9 × 9 in. photograph in two multiple sections (e.g., 8.5 × 9 in. and 0.5 × 9 in. for a 9 × 9 in. aerial photograph) and then digitally *mosaic* the two pieces together. The mosaicking process can introduce both geometric and radiometric error. Therefore, it is better to use as high-quality and as large a digitizer as possible to minimize

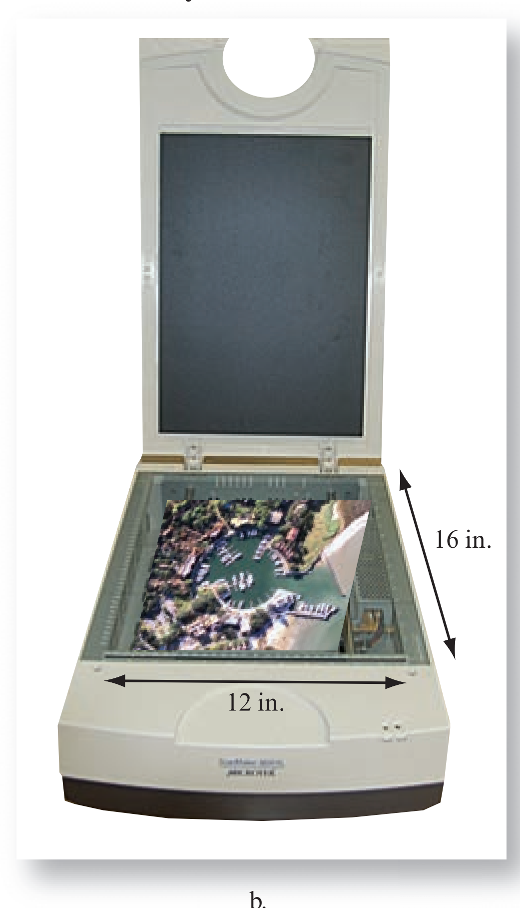
Linear Array CCD

|◄— 3,000 photosites —►|

a.

Area Array CCD

c.



b.

FIGURE 2–4 a) A linear charge-coupled device (CCD) containing 3,000 photosites. b) An oversized flatbed scanner based on linear array CCD technology capable of digitizing hard-copy maps and aerial photographs that are up to 12 × 16 in. It is very useful for digitizing 9 × 9 in. aerial photographs in a single pass. c) An example of an area array CCD used in an area-array scanner (from Jensen and Jensen, 2013).

the amount of mosaicking required. A high-quality 12 × 16 in. scanner is shown in Figure 2-4b.

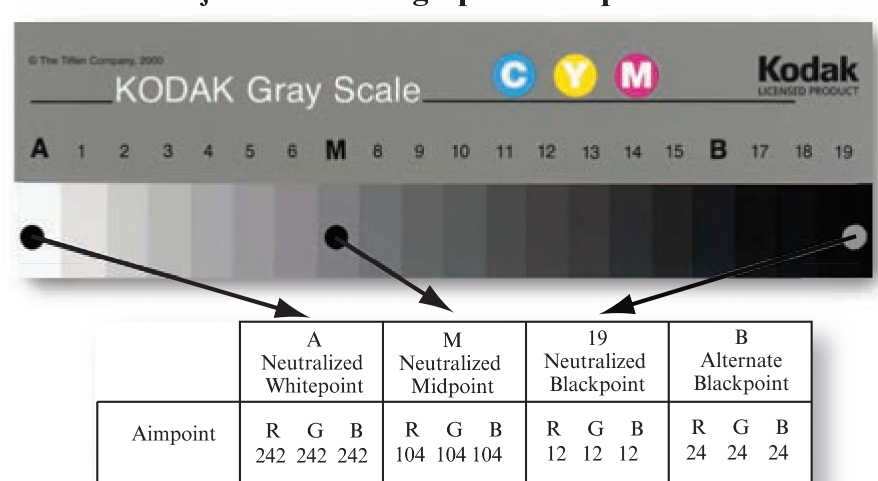
Some digitizing systems utilize area array CCD technology (Figure 2-4c). These systems scan the film (the original negative or positive transparency) as a series of rectangular image segments or tiles. Radiometric calibration algorithms are then used to compensate for uneven illumination encountered in any of the tile regions. When scanning a color image, the scanner stops on a rectangular image section and captures that information sequentially with each of the color filters (blue, green, red) before it moves to another section.

When scanning black-and-white or color images, it is good practice to use special Gray Scale and Color Control Patches such as those shown in Figure 2-5ab. The U. S. National Archives and Records Administration (NARA) recommends that these cards be laid next to the map or image that is to be digitized (Puglia et al., 2004). After digitizing, the analyst looks at the quality of the red, green, and blue (RGB) values of the white, gray, and black parts of the grayscale or color control patches on the computer screen. If the values of the RGB values for the white, gray, and black test areas lie within the "aimpoint" range, then it is likely that the digitization has been successful. If the values fall outside the aimpoint range, then adjustments should be made and the image or map should be redigitized until aimpoint values are achieved.

Digitized National Aerial Photography Program (NAPP) Data

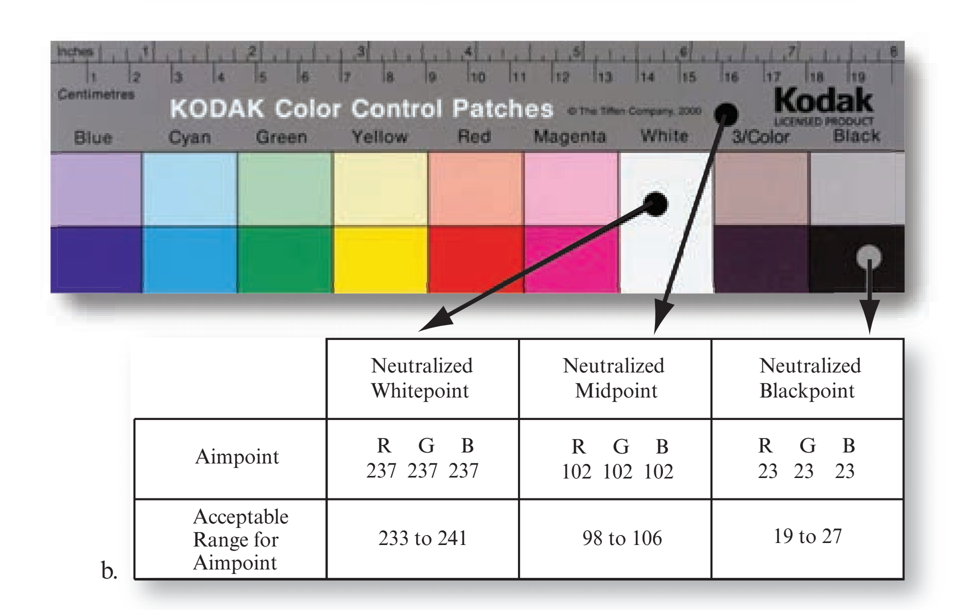
The National Aerial Photography Program (NAPP) was initiated in 1987 as a replacement for the National High Altitude Aerial Photography (NHAP) Program. The objective of the National Aerial Photography Program is to acquire and archive photographic coverage of the coterminous United States at 1:40,000 scale using color-infrared or black-and-white film. The photography is acquired at an altitude of 20,000 ft. above

Gray Scale and Color Control Patches
Placed Adjacent to Photographs or Maps to be Scanned



	A Neutralized Whitepoint	M Neutralized Midpoint	19 Neutralized Blackpoint	B Alternate Blackpoint
Aimpoint	R G B 242 242 242	R G B 104 104 104	R G B 12 12 12	R G B 24 24 24
Acceptable Range for Aimpoint	239 to 247	100 to 108	8 to 16	20 to 28

a.



	Neutralized Whitepoint	Neutralized Midpoint	Neutralized Blackpoint
Aimpoint	R G B 237 237 237	R G B 102 102 102	R G B 23 23 23
Acceptable Range for Aimpoint	233 to 241	98 to 106	19 to 27

b.

FIGURE 2–5 a) A Gray Scale card is placed adjacent to a black-and-white aerial photograph to be scanned. The goal is to have the RGB values in the digital output file match as close as possible the aimpoint RGB values specified in the table for white, mid, and black points. b) The Color Control Patch card is placed adjacent to the color aerial photograph to be scanned (courtesy of U.S. National Archives and Records Administration; Puglia et al., 2004; Jensen and Jensen, 2013).

ground level (AGL) with a 6-in. focal-length metric camera. The aerial photography is acquired ideally on a five-year cycle, resulting in a nationwide photographic database that is readily available through the EROS Data Center in Sioux Falls, SD, or the Aerial Photography Field Office in Salt Lake City, UT.

High spatial resolution NAPP photography represents a wealth of information for on-screen photo interpretation and can become a high-resolution basemap upon which other GIS information (e.g., parcel boundaries, utility lines, tax data) may be overlaid after it is digitized and rectified to a standard map projection. Light (1993) summarized the optimum methods for converting the NAPP data into a national database of digitized photography that meets National Map Accuracy Standards. Microdensitometer scanning of the photography, using a spot size of 15 μm, preserves the 27 resolvable line-pair-per-millimeter (lp/mm) spatial resolution in the original NAPP photography. This process generally yields a digital dataset that has a ground spatial resolution of 1 × 1 m, depending on

original scene contrast. This meets most land-cover and land-use mapping user requirements.

The digitized information can be color-separated into separate bands of information if desired. The 15-μm scanning spot size will support most digital soft-copy photogrammetry for which coordinate measurements are made using a computer and the monitor screen (Light, 1993). Because the digitized NAPP data are so useful as a high spatial resolution GIS basemap, many states have entered cost-sharing relationships with the U.S. Geological Survey and have their NAPP coverage digitized and output as digital orthophotomaps. A large amount of NAPP data has been digitized and converted into digital orthophotoquads (Light, 1996).

Digitization Considerations

There are some basic guidelines associated with digitizing aerial photography or other types of hard-copy remote sensor data. First, the person who will be digitally processing the digitized remote sensor data should make the decision about what dpi to use (200 dpi, 1,000 dpi, etc.) based on the scale of the original imagery and the desired spatial resolution (e.g., 1:40,000-scale aerial photography scanned at 1,000 dpi yields 1 × 1 m spatial resolution pixels; refer to Table 2-1). For example, consider the digitized NAPP photography of Three Mile Island shown in Figure 2-6. The original 1:40,000-scale panchromatic photography was scanned at 25 to 1,000 dpi. The U.S. Geological Survey typically scans NAPP photography at 1,000 dpi to produce digital orthophotoquads. Note that for all practical purposes it is difficult to *discern visually on the printed page* any significant difference in the quality of the digitized data until about 150 dpi. Then, the information content gradually deteriorates until at 72 to 25 dpi, individual features in the scene are very difficult to interpret. It is important to remember, however, that when the various digitized datasets are displayed on a CRT screen and magnification takes place (as simulated in the bottom nine images in Figure 2-6), it is quite evident that there is significantly more information content in the higher-resolution scans (e.g., 1,000 dpi) than the lower-resolution scans (e.g., 100 dpi).

A second general principle is that when digitizing large-scale imagery, it is not necessary to scan at extremely high rates (e.g., 300 – 1,000 dpi) to obtain visually acceptable imagery. For example, large-scale helicopter vertical photography of vehicles in a parking lot in Waikiki, HI, was digitized using scan resolutions from 10 to 1,000 dpi (Figure 2-7). Once again, there is very little noticeable difference in the quality of the information at the high scan resolutions (e.g., 1,000, 500, 300, and 200). Note however, that even the 100-dpi and possibly the 72-dpi digitized images still contain significant detailed information *on the printed page* because the original aerial photography was obtained at a very large scale.

Hopefully, there is a relationship between the brightness value ($BV_{i,j,k}$) or density ($D_{i,j}$) at any particular location in the digitized image and the energy reflected from the real-world object space ($O_{x,y}$) at the exact location. Scientists take advantage of this relationship by 1) making careful *in situ* observations in the field, such as the amount of biomass for a 1 × 1 m spot on the Earth located at $O_{x,y}$, and then 2) measuring the brightness value ($BV_{i,j,k}$) or density of the object at that exact location in the photograph using a *densitometer*. If enough samples are located in the field and in the photography, it may be possible to develop a correlation between the real-world object space and the image space. This is an important use of digitized aerial photography.



Digital Remote Sensor Data Collection

The previous section was devoted to the digitization of aerial photography or other types of hard-copy remotely sensed data. Digitized natural color and color-infrared aerial photography can be considered three-band multispectral datasets. Properly digitized natural color aerial photography can be converted to blue, green, and red bands of registered digital data. Digitized color-infrared aerial photography can be converted to green, red, and near-infrared bands of digital data. Although these three-band multispectral datasets are sufficient for many applications, there are times when even more spectral bands located at optimum locations throughout the electromagnetic spectrum would be useful for a specific application. Fortunately, optical engineers have developed detectors that are sensitive to hundreds of bands in the electromagnetic spectrum. The measurements made by the detectors are usually stored in a digital format.

Multispectral remote sensing is defined as the collection of reflected, emitted, or back-scattered energy from an object or area of interest in multiple bands (regions) of the electromagnetic spectrum. **Hyperspectral remote sensing** involves data collection in hundreds of bands. **Ultraspectral remote sensing** involves data collection in many hundreds of bands. Most multispectral and hyperspectral remote sensing systems collect data in a digital format. The remainder of this chapter introduces the characteristics of historical, current, and proposed multispectral and hyperspectral remote sensing systems.

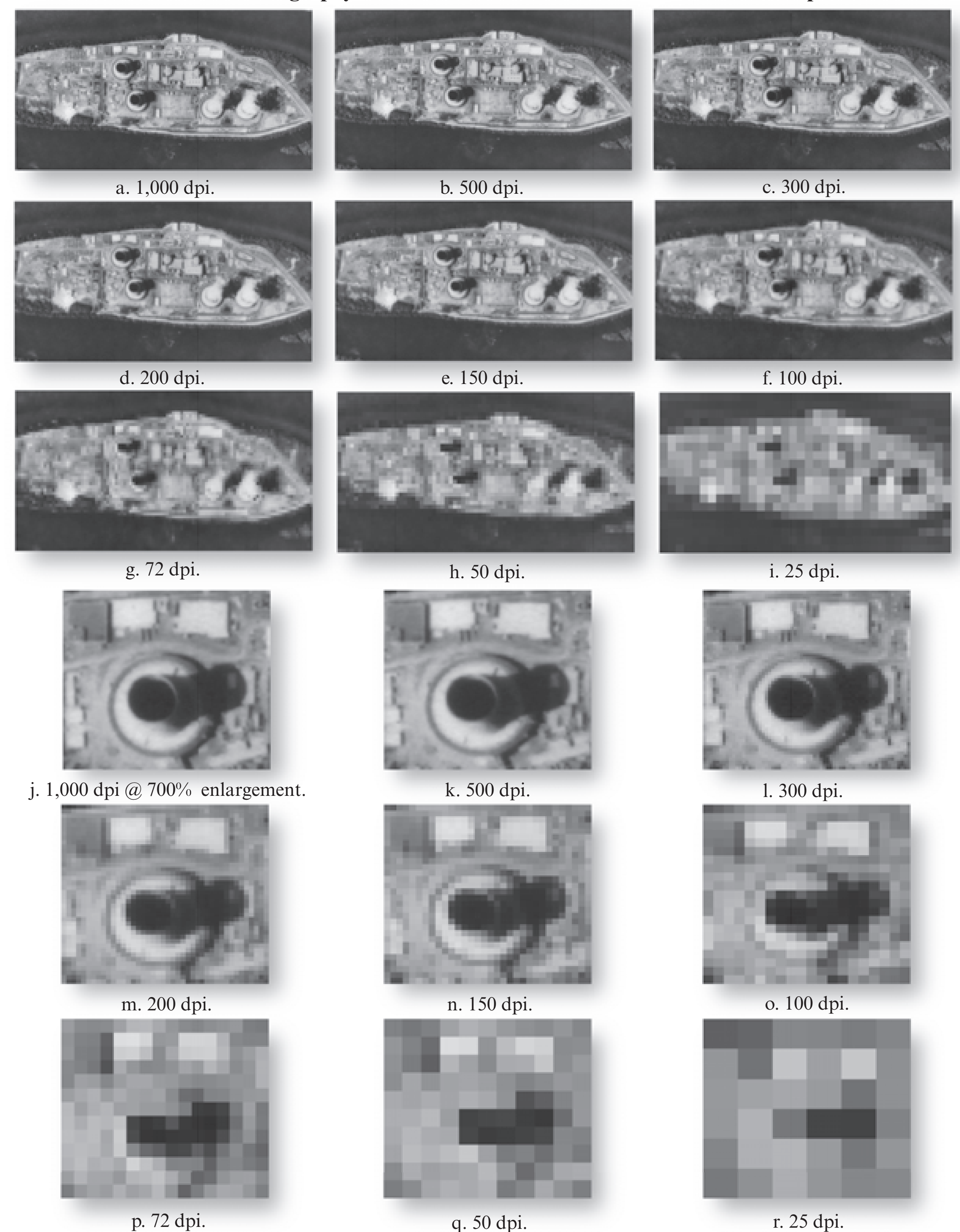


FIGURE 2–6 National Aerial Photography Program (NAPP) photography of Three Mile Island, PA, digitized at various resolutions from 1,000 to 25 dpi. The original 9 × 9 in. vertical aerial photograph was obtained on September 4, 1987, at an altitude of 20,000 ft. above ground level yielding a scale of 1:40,000. Scanning at 1,000 dpi yields 1 × 1 m pixels (refer to Table 2-1). The upper nine images are printed at approximate contact scale while the lower nine are enlarged to demonstrate approximate information content (NAPP photography courtesy of the U.S. Geological Survey).

Large-scale Vertical Aerial Photography Scanned at Various Dots-per-inch

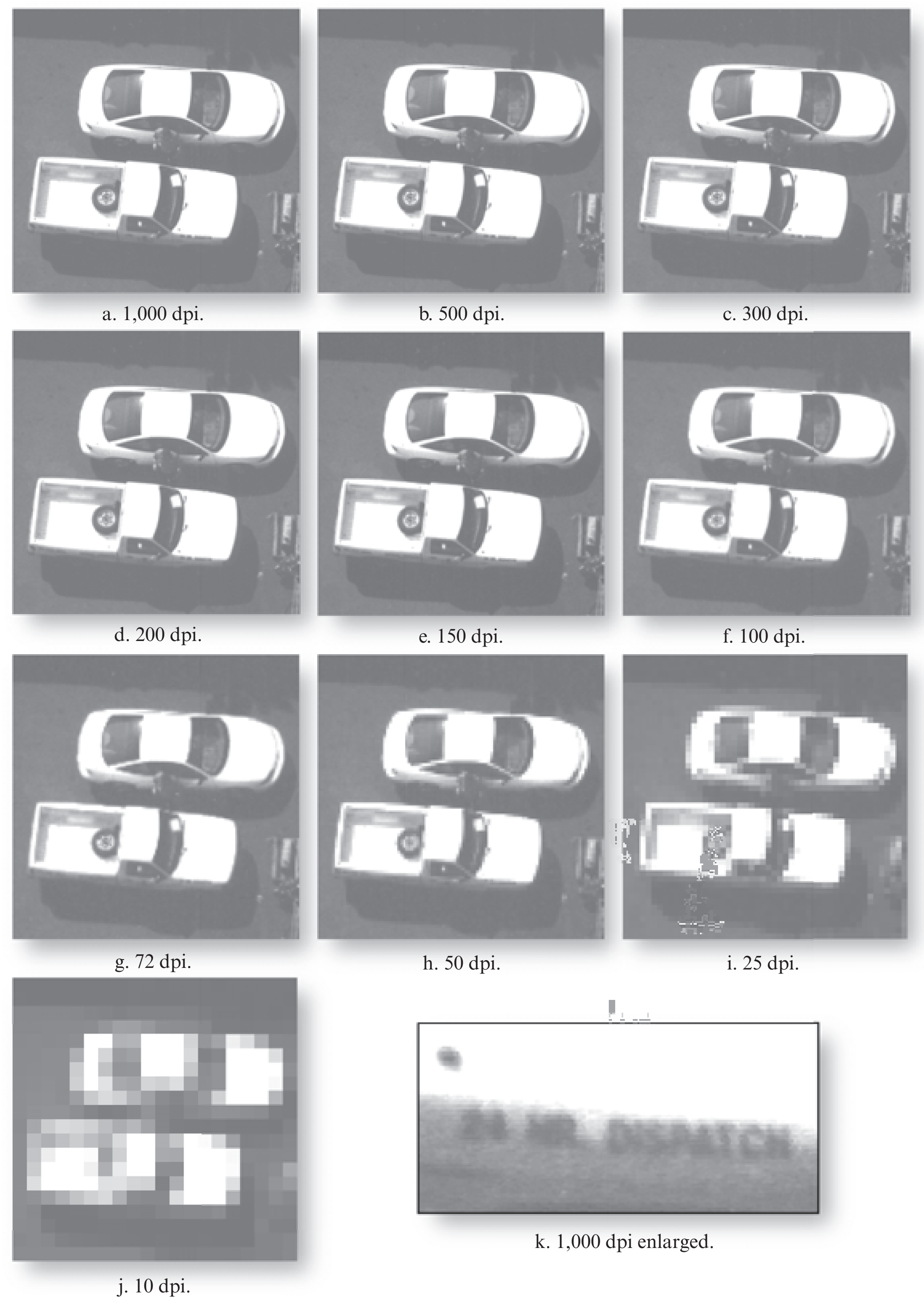


FIGURE 2–7 Digitized large-scale panchromatic aerial photography of vehicles in a parking lot in Waikiki, HI. The original photograph was obtained from a helicopter platform and then scanned at rates from 1,000 to 10 dpi. It is possible to make out the words "24 HR DISPATCH" on the vehicle when the photography is scanned at 1,000 dpi and magnified.

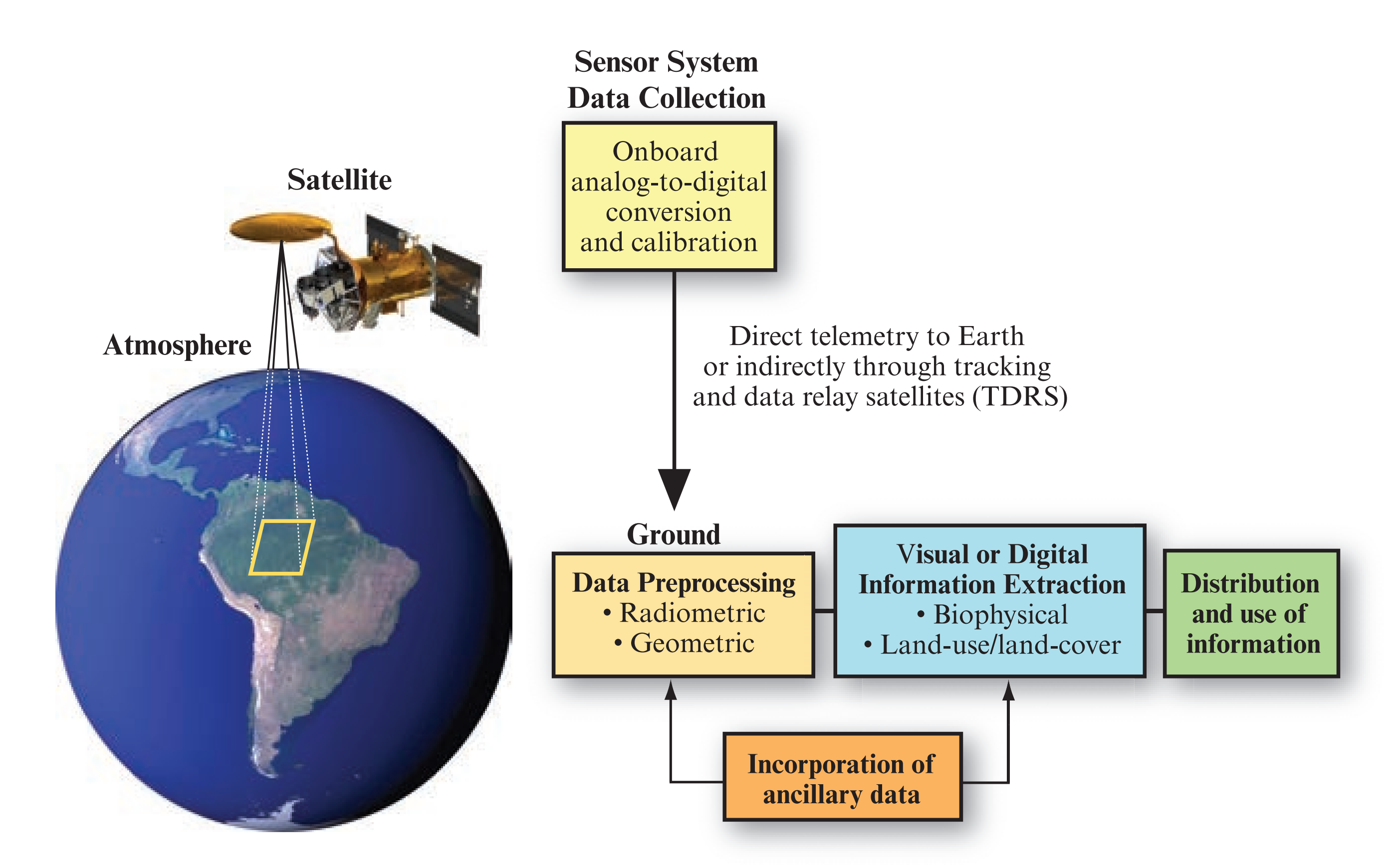


FIGURE 2–8 An overview of the way digital remotely sensed data are transformed into useful information. The data recorded by the detectors are often converted from an analog electrical signal to a digital value and calibrated. Ground preprocessing removes geometric and radiometric distortions. This may involve the use of ephemeris or ancillary (collateral) data such as map *x*, *y* coordinates, a digital elevation model, etc. The data are then ready for visual or digital analysis to extract biophysical or land-use/land-cover information. Future sensor systems will conduct preprocessing and information extraction onboard the remote sensing system (*Blue Marble* image of the Earth courtesy of NASA Earth Observatory).

An overview of how digital remote sensor data are turned into useful information is shown in Figure 2-8. The remote sensor system first detects electromagnetic energy that exits from the phenomena of interest and passes through the atmosphere. The energy detected is recorded as an analog electrical signal, which is usually converted into a digital value through an A-to-D conversion. If an aircraft platform is used, the digital data are simply returned to Earth. However, if a spacecraft platform is used, the digital data are telemetered to Earth receiving stations directly or indirectly via tracking and data relay satellites (TDRS). In either case, it may be necessary to perform some radiometric or geometric preprocessing of the digital remotely sensed data to improve their interpretability. The data can then be enhanced for subsequent human visual analysis or processed further using digital image processing algorithms. Biophysical, land-use, or land-cover information extracted using either a visual or digital image processing approach is distributed and used to make decisions.

There are a great variety of digital multispectral and hyperspectral remote sensing systems. It is beyond the scope of this book to provide detailed information on each of them. However, it is possible to review selected remote sensing systems that are or will be of significant value for Earth resource investigations. They are organized according to the type of remote sensing technology used, as summarized in Figure 2-9, including:

Multispectral Imaging Using Discrete Detectors and Scanning Mirrors

- Landsat Multispectral Scanner (MSS)
- Landsat Thematic Mapper (TM)
- Landsat 7 Enhanced Thematic Mapper Plus (ETM^+)
- NOAA Geostationary Operational Environmental Satellite (GOES)
- NOAA Advanced Very High Resolution Radiometer (AVHRR)
- NPOESS Preparatory Project (NPP)

Multispectral Imaging Using Linear Arrays

- Landsat 8 (linear array "pushbroom")
- SPOT Image, Inc. (SPOT 1-7 and *Vegetation* sensor)
- European Space Agency (Pleiades 1A and 1B)
- Indian Remote Sensing System (IRS, CartoSat, ResourceSat)
- Korean Research Institute (KOMPSAT)
- Astrium, Inc. (Sentinel-2)
- NASA *Terra* Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)
- NASA *Terra* Multi-angle Imaging Spectroradiometer (MISR)
- Space Imaging/GeoEye, Inc. (IKONOS, GeoEye-1)
- EarthWatch/DigitalGlobe, Inc. (QuickBird, WorldView-1, -2, -3)
- ImageSat International, Inc. (EROS A1 and A2)
- RapidEye, Inc.

Imaging Spectrometry Using Linear and Area Arrays

- NASA Earth Observer (EO-1) Advanced Land Imager (ALI), Hyperion, and LEISA Atmospheric Corrector (LAC)
- NASA Jet Propulsion Laboratory Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)
- Itres, Inc., Compact Airborne Spectrographic Imager (CASI-1500)
- NASA *Terra* Moderate Resolution Imaging Spectrometer (MODIS)
- NASA HyspIRI

Airborne Digital Cameras

- UltraCAM
- Leica

Satellite Analog and Digital Photographic Systems

- NASA Space Shuttle and International Space Station Imagery



Multispectral Imaging Using Discrete Detectors and Scanning Mirrors

The collection of multispectral remote sensor data using discrete detectors and scanning mirrors has been with us since the mid-1960s. Despite the technology's age, several new remote sensing systems still use it.

Earth Resource Technology Satellites and Landsat 1-7 Sensor Systems

In 1967, the National Aeronautics & Space Administration (NASA), encouraged by the U.S. Department of the Interior, initiated the Earth Resource Technology Satellite (ERTS) program. This program resulted in the deployment of five satellites carrying a variety of remote sensing systems designed primarily to acquire Earth resource information. The most noteworthy sensors were the Landsat Multispectral Scanner and the Landsat Thematic Mapper (Table 2-2). The Landsat program is the United States' oldest land-surface observation satellite system, having obtained data since 1972. It has had a tumultuous history of management and funding sources.

The chronological launch and retirement history of the satellites is shown in Figure 2-10. The ERTS-1 satellite, launched on July 23, 1972, was the first experimental system designed to test the feasibility of collecting Earth resource data by unmanned satellites. Prior to the launch of ERTS-B on January 22, 1975, NASA renamed the ERTS program *Landsat*, distinguishing it from the *Seasat* oceanographic RADAR satellite launched on June 26, 1978. At this time, ERTS-1 was retroactively named Landsat 1 and ERTS-B became Landsat 2 at launch. Landsat 3 was launched March 5, 1978; Landsat 4 on July 16, 1982; and Landsat 5 on March 1, 1984.

The Earth Observation Satellite Company (EOSAT) was given control of the Landsat satellites in September, 1985. Unfortunately, Landsat 6 with its Enhanced Thematic Mapper (ETM) (a 15 × 15 m panchromatic band was added) failed to achieve orbit on October 5, 1993. Landsat 7 with its Enhanced Thematic Mapper Plus (ETM^+) sensor system was launched on April 15, 1999. For a detailed history of the Landsat program, refer to the *Landsat Data User Notes* published by the EROS Data Center; *Imaging Notes* published by Space Imaging, Inc.; the NASA Landsat 7 home page (NASA Landsat 7, 2014); Irons et al. (2012); and Loveland and Dwyer (2012).

Landsats 1 through 3 were launched into circular orbits around Earth at a nominal altitude of 919 km (570 mi). The platform is shown in Figure 2-11a. The satellites had an orbital inclination of 99°, which made them nearly polar (Figure 2-11b) and caused them to cross the equator at an angle of approximately 9° from normal. The satellites orbited Earth once every 103 minutes, resulting in 14 orbits per day (Figure 2-11c). This Sun-synchronous orbit meant that the orbital plane precessed around Earth at the same angular rate at which Earth moved around the Sun. This characteristic caused the satellites to cross the equator at the

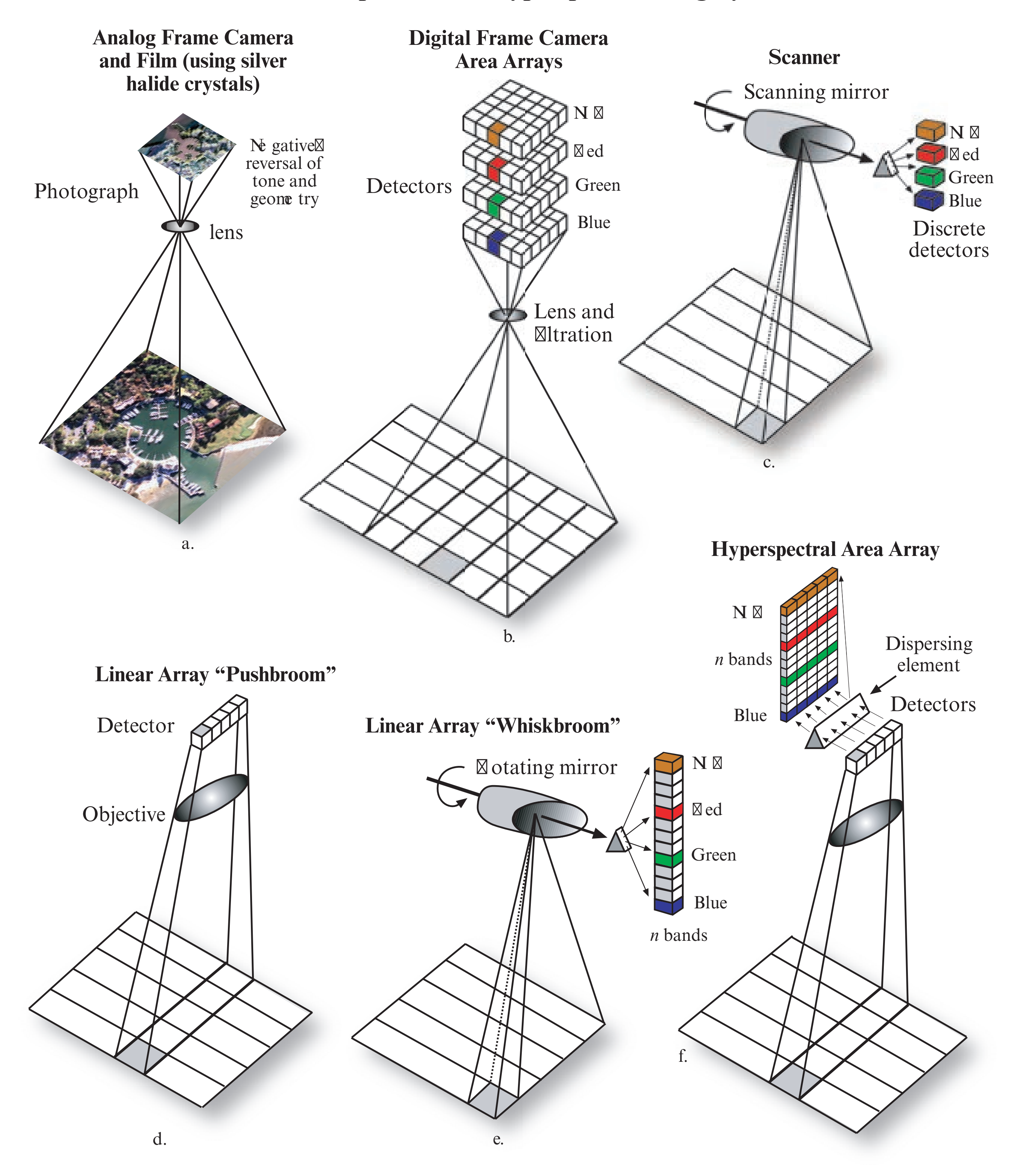


FIGURE 2–9 Six types of remote sensing systems used for multispectral and hyperspectral data collection: a) traditional analog (film) aerial photography, b) digital frame camera aerial photography based on area arrays, c) imaging using a scanning mirror and discrete detectors, d) multispectral imaging using linear arrays (often referred to as "pushbroom" technology), e) imaging with a scanning mirror and linear arrays (often referred to as "whiskbroom" technology), and f) imaging spectrometry using linear and area arrays.

TABLE 2–2 Landsat Multispectral Scanner (MSS) and Landsat Thematic Mapper (TM) 4 and 5 sensor system characteristics.

	Landsat Multispectral Scanner (MSS)		Landsat 4 and 5 Thematic Mapper (TM)	
	Band	**Spectral Resolution (μm)**	**Band**	**Spectral Resolution (μm)**
	4[a] Green	0.5 – 0.6	1 Blue	0.45 – 0.52
	5 Red	0.6 – 0.7	2 Green	0.52 – 0.60
	6 Near-infrared	0.7 – 0.8	3 Red	0.63 – 0.69
	7 Near-infrared	0.8 – 1.1	4 Near-infrared	0.76 – 0.90
	8[b] Thermal Infrared	10.4 – 12.6	5 SWIR	1.55 – 1.75
			6 Thermal Infrared	10.40 – 12.5
			7 SWIR	2.08 – 2.35
IFOV at nadir	79 × 79 m for bands 4 through 7 240 × 240 m for band 8		30 × 30 m for bands 1 through 5, 7 120 × 120 m for band 6	
Data rate	15 Mb/s		85 Mb/s	
Quantization levels	6 bit (values from 0 to 63)		8 bit (values from 0 to 255)	
Earth coverage	18 days Landsat 1, 2, 3 16 days Landsat 4, 5		16 days Landsat 4, 5	
Altitude	919 km		705 km	
Swath width	185 km		185 km	
Inclination	99°		98.2°	

a. MSS bands 4, 5, 6, and 7 were renumbered bands 1, 2, 3, and 4 on Landsats 4 and 5.
b. MSS band 8 was present only on Landsat 3.

same local time (9:30 to 10:00 a.m.) on the illuminated side of Earth.

Figures 2-11c and 2-12 illustrate how repeat coverage of a geographic area was acquired. From one orbit to the next, a position directly below the spacecraft moved 2,875 km (1,785 mi) at the equator as the Earth rotated beneath it. The next day, 14 orbits later, it was back to its original location, with orbit 15 displaced westward from orbit 1 by 159 km (99 mi) at the equator. This continued for 18 days, after which orbit 252 fell directly over orbit 1 once again. Landsat had the capability of observing the entire globe (except poleward of 81°) once every 18 days, or about 20 times a year. There were approximately 26 km (16 mi) of overlap between successive orbits. This overlap was a maximum at 81° North and South latitudes (about 85%) and a minimum at the equator (about 14%). This has proven useful for stereoscopic analysis applications.

An elegant method of determining if remote sensor data (e.g., Landsat MSS, Thematic Mapper, ETM^{+}) are available for a specific location is to use the U.S. Geological Survey's *Global Visualization Viewer* (USGS GloVis, 2014). For example, suppose we are interested in locating a Landsat Thematic Mapper image of Charleston, SC. We can enter the Global Visualization Viewer and specify WRS Path 16 and Row 37 and search the database as shown in Figure 2-13a. If we do not know the path and row designation, we could a) move the cursor on the regional map and place it on Charleston, SC, or b) input the latitude and longitude coordinates of Charleston, SC (33.2°N, -81°W). We can specify the month and year (e.g., March, 2010), and the amount of acceptable cloud cover (e.g., 0%). We can also specify whether the search should be conducted regionally as in Figure 2-13a (pixels resampled to 1,000 m) or locally as in Figure 2-13b (pixels resampled to 240 m).

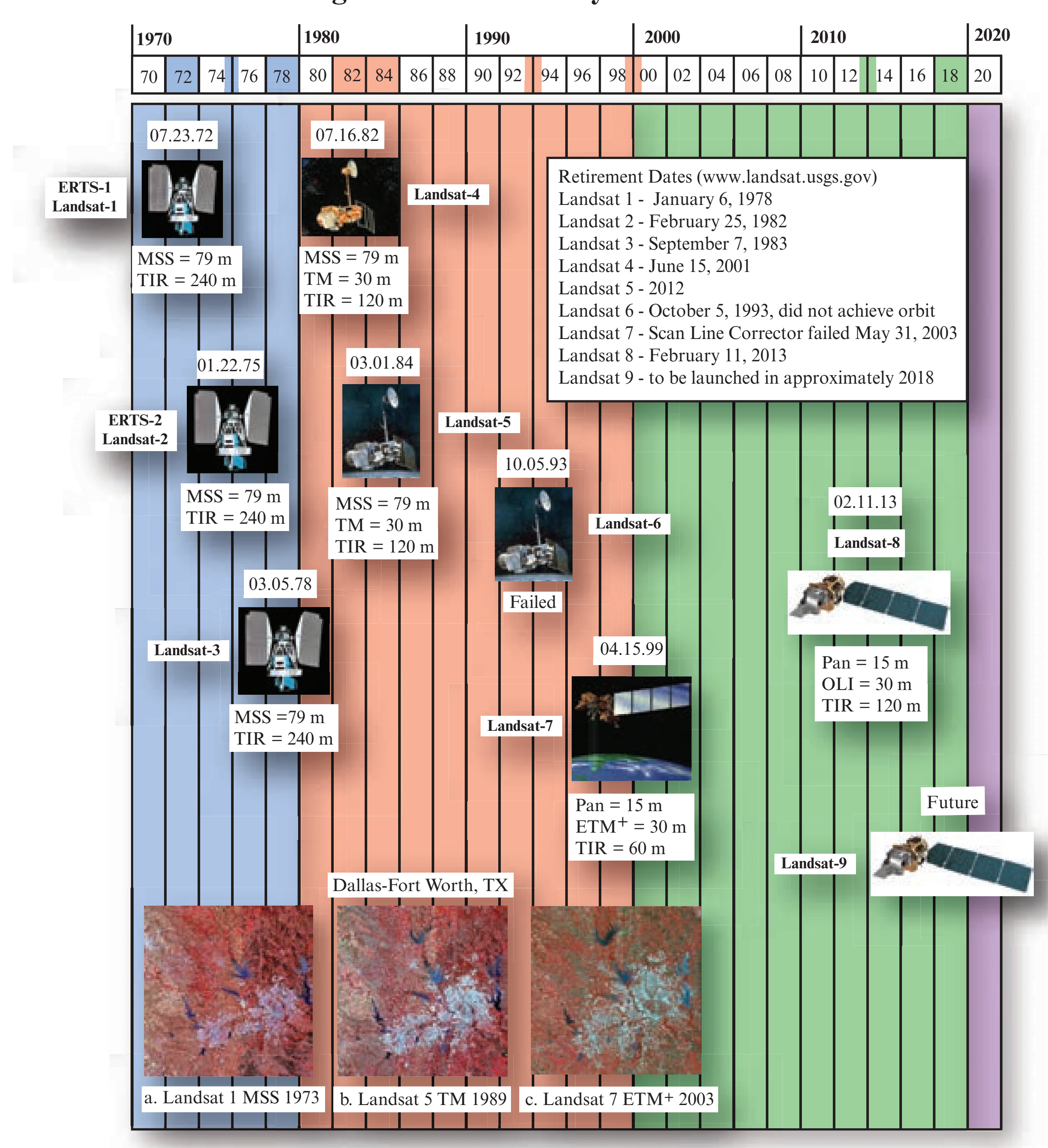
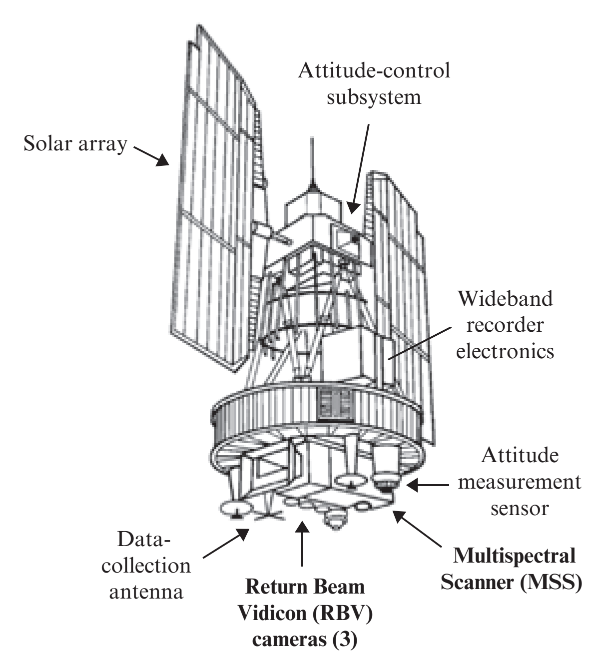


FIGURE 2–10 Chronological launch, retirement history, and proposed Landsat satellites. Multiple-date Landsat images of Dallas-Fort Worth, TX, are provided (updated from Jensen et al., 2012).

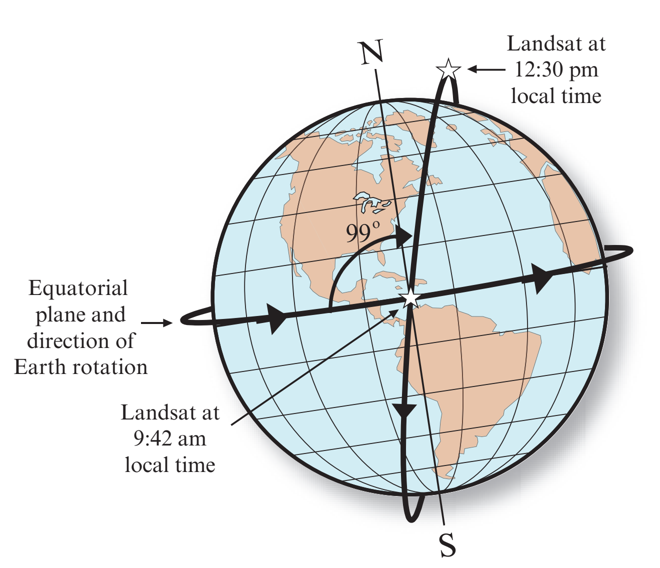
In the context of this section on data acquisition, we are interested in the type of sensors carried aloft by the Landsat satellites and the nature and quality of remote sensor data provided for Earth resource investigations. The most important sensors were the Multispectral Scanner and Landsat Thematic Mapper.

Landsat Multispectral Scanner

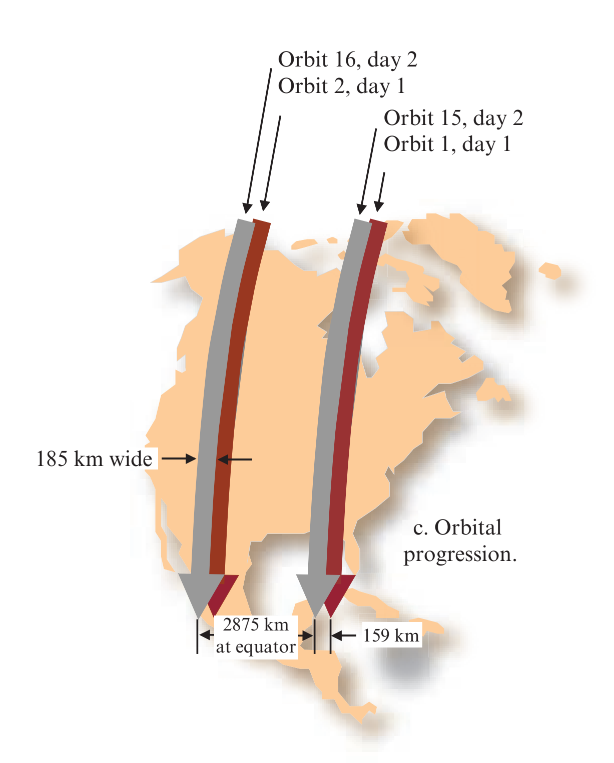
The Landsat Multispectral Scanner (MSS) was placed on Landsat satellites 1 through 5. The MSS multiple-detector array and the scanning system are shown diagrammatically in Figure 2-13a. Sensors such as the Landsat MSS (and Thematic Mapper to be discussed) are optical-mechanical systems in which a mirror scans



a. Landsat MSS 1, 2, and 3 Nimbus-style platform.



b. Sun-synchronous orbital inclination.



c. Orbital progression.

FIGURE 2–11 a) Nimbus-style platform used for Landsats 1, 2, and 3 and associated sensor and telecommunication systems. b) Inclination of the Landsat orbit to maintain a Sun-synchronous orbit. c) From one orbit to the next, the position directly below the satellite moved 2,875 km (1,785 mi) at the equator as Earth rotated beneath it. The next day, 14 orbits later, it was approximately back to its original location, with orbit 15 displaced westward from orbit 1 by 159 km (99 mi). This is how repeat coverage of the same geographic area was obtained.

the terrain perpendicular to the flight direction. While it scanned, it focused energy reflected or emitted from the terrain onto discrete detector elements. The detectors converted the radiant flux measured within each instantaneous-field-of-view (IFOV) in the scene into an electronic signal (Figure 2-13a). The detector elements were placed behind filters that pass broad portions of the spectrum. The MSS had four sets of filters and detectors, whereas the TM had seven. The primary limitation of this approach was the short residence time of the detector in each IFOV. To achieve adequate signal-to-noise ratio without sacrificing spatial resolution, such a sensor had to operate in broad spectral bands of ≥ 100 nm or use optics with unrealistically small ratios of focal-length to aperture (*f* / stop).

The MSS scanning mirror oscillated through an angular displacement of ±5.78° off-nadir. This 11.56° field of view resulted in a swath width of approximately 185 km (115 mi) for each orbit. Six parallel detectors sensitive to four spectral bands (channels) in the electromagnetic spectrum viewed the ground simultaneously:

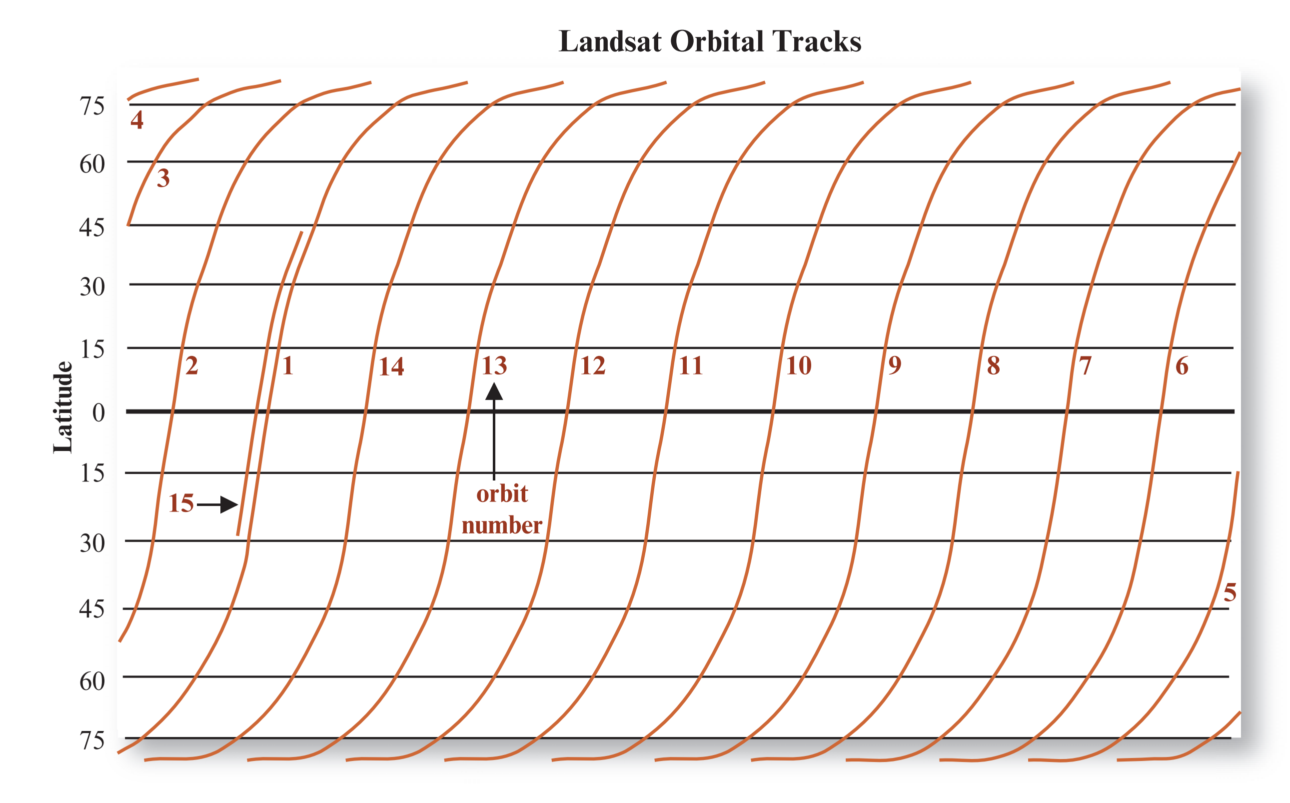


FIGURE 2–12 Orbital tracks of Landsat 1, 2, or 3 during a single day of coverage. The satellite crossed the equator every 103 minutes, during which time the Earth rotated a distance of 2,875 km under the satellite at the equator. Every 14 orbits, 24 hours elapsed.

0.5 – 0.6 μm (green), 0.6 – 0.7 μm (red), 0.7 – 0.8 μm (reflective infrared), and 0.8 – 1.1 μm (reflective infrared). These bands were originally numbered 4, 5, 6, and 7, respectively, because a Return-Beam-Vidicon (RBV) sensor system also onboard the satellite recorded energy in three bands labeled 1, 2, and 3.

When not viewing the Earth, the MSS detectors were exposed to internal light and Sun calibration sources. The spectral sensitivity of the bands is summarized in Table 2-2 and shown diagrammatically in Figure 2-14b. Note that there is spectral overlap between the bands.

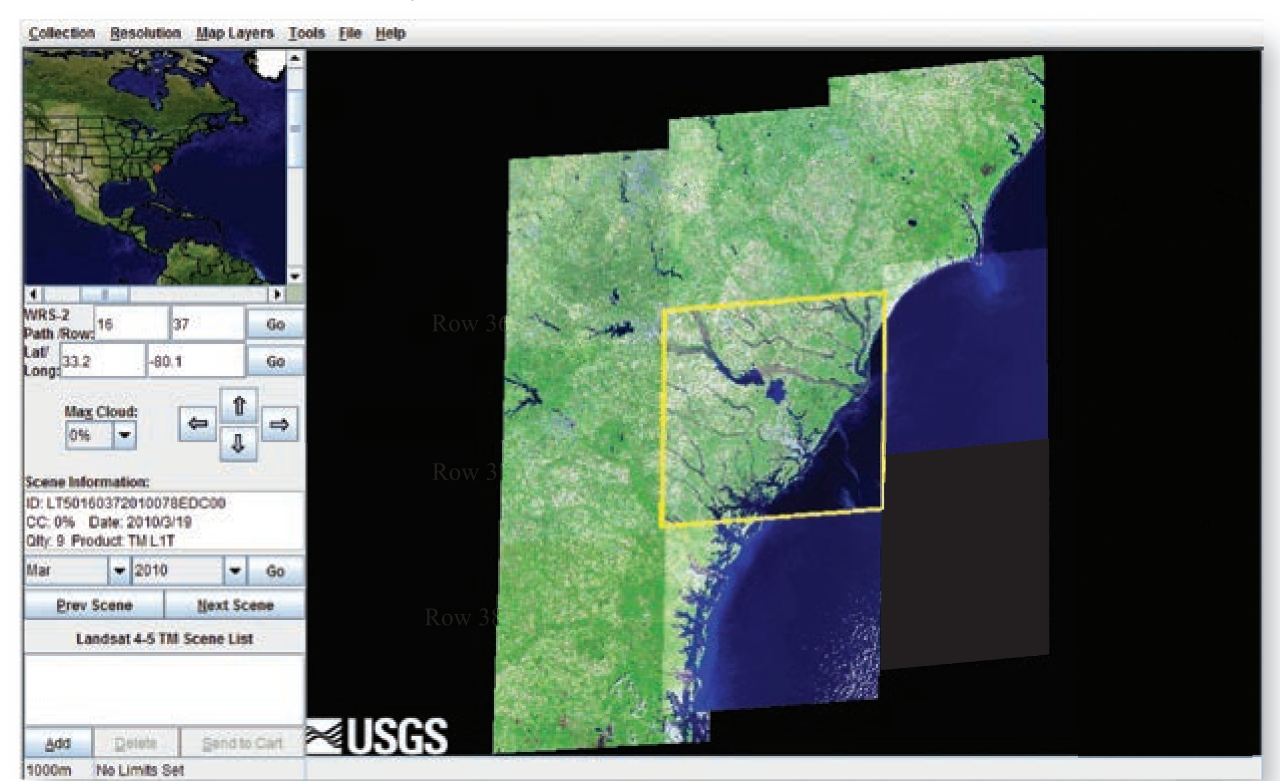
The IFOV of each detector was square and resulted in a ground resolution element of approximately 79 × 79 m (67,143 ft^2). The voltage analog signal from each detector was converted to a digital value using an onboard A-to-D converter. The data were quantized to 6 bits with a range of values from 0 to 63. These data were then rescaled to 7 bits (0 to 127) for three of the four bands in subsequent ground processing (i.e., bands 4, 5, and 6 were decompressed to a range of 0 to 127). It is important to remember that the early 1970s Landsat MSS data were quantized to 6-bits when comparing MSS data collected in the late 1970s and 1980s, which were collected at 8-bits.

The MSS scanned each line across-track from west to east as the southward orbit of the spacecraft provided the along-track progression. Each MSS scene represents a 185 × 170 km parallelogram extracted from the continuous swath of an orbit and contains approximately 10 percent overlap. A typical scene contains approximately 2,340 scan lines with about 3,240 pixels per line, or about 7,581,600 pixels per channel. All four bands represent a dataset of more than 30 million brightness values. Landsat MSS images provided an unprecedented ability to observe large geographic areas while viewing a single image. For example, approximately 5,000 conventional vertical aerial photographs obtained at a scale of l:15,000 are required to equal the geographic coverage of a single Landsat MSS image. This allows regional terrain analysis to be performed using one data source rather than a multitude of aerial photographs.

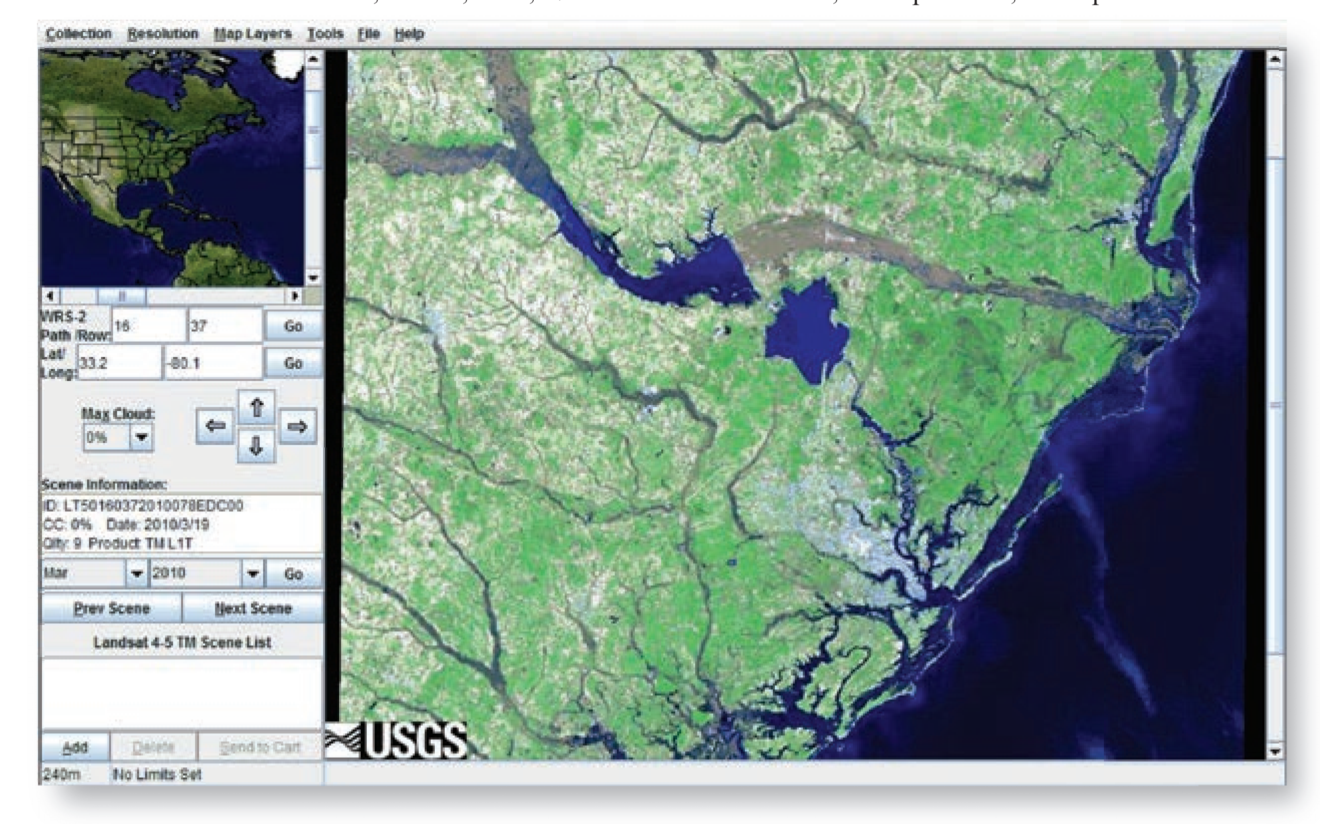
Landsat Thematic Mapper (TM)

Landsat Thematic Mapper sensor systems were launched on July 16, 1982 (Landsat 4), and on March 1, 1984 (Landsat 5). The TM is an optical-mechanical "whiskbroom" sensor that records energy in the visible, reflective-infrared, middle-infrared, and thermal-infrared regions of the electromagnetic spectrum. It collects multispectral imagery that has higher spatial, spectral, temporal, and radiometric resolution than the

USGS *Global Visualization Viewer* used to Locate Landsat Remote Sensor Data



a. Search for Landsat 4-5 TM images of Charleston, SC, using the criteria:
Path 16 and Row 37; March, 2010; 0% maximum cloud cover; resampled to 1,000 m pixels.



b. Search for Landsat 4-5 TM imagery of Charleston, SC, resampled to 240 m pixels.

FIGURE 2–13 Use of the USGS *Global Visualization Viewer* to locate Landsat Thematic Mapper imagery of Charleston, SC.

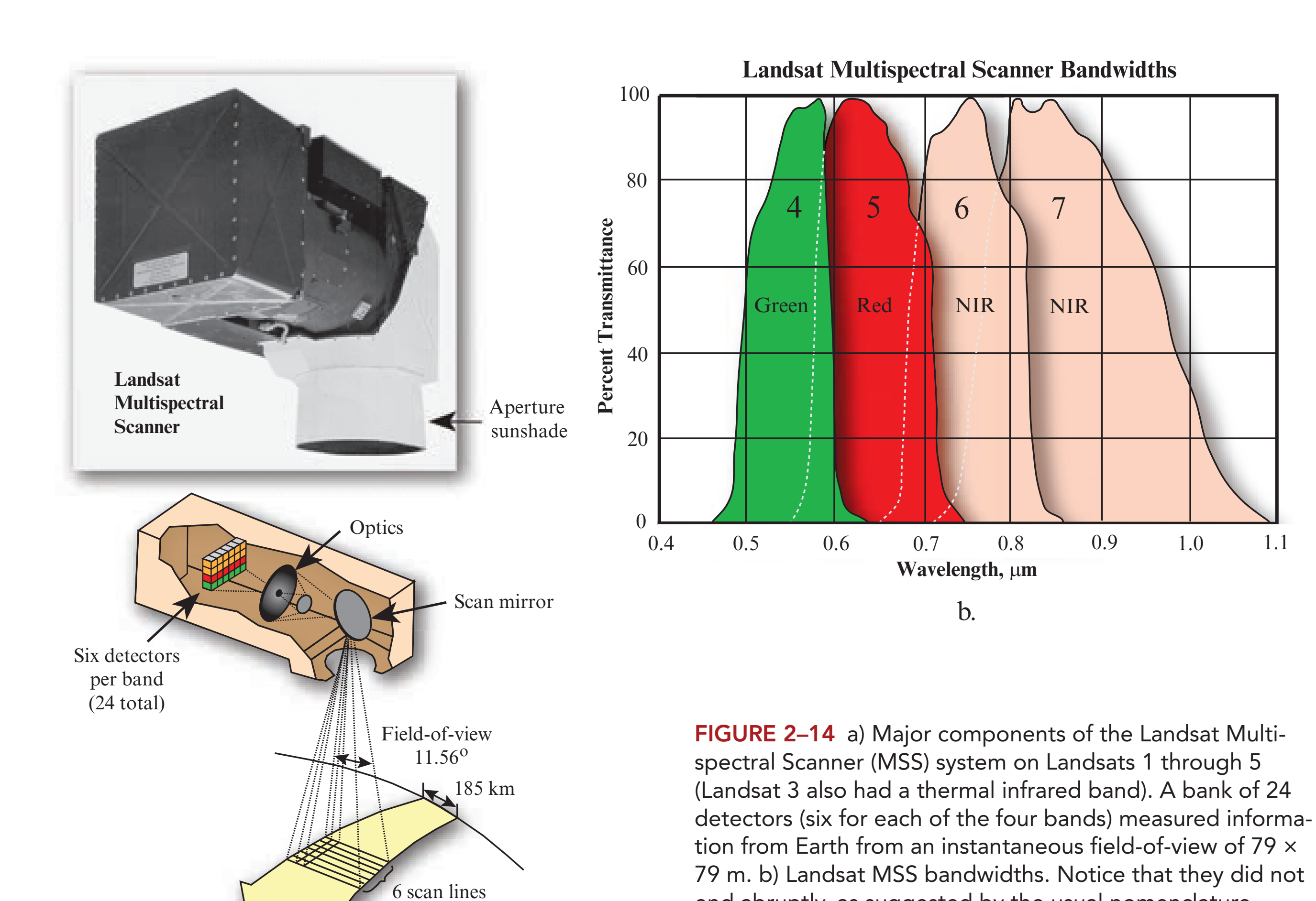


FIGURE 2–14 a) Major components of the Landsat Multispectral Scanner (MSS) system on Landsats 1 through 5 (Landsat 3 also had a thermal infrared band). A bank of 24 detectors (six for each of the four bands) measured information from Earth from an instantaneous field-of-view of 79 × 79 m. b) Landsat MSS bandwidths. Notice that they did not end abruptly, as suggested by the usual nomenclature.

Landsat MSS. Detailed descriptions of the design and performance characteristics of the TM can be found at NASA Landsat 7 (2014).

The Landsat 4 and 5 platform and sensor are shown in Figure 2-15. The Thematic Mapper sensor system configuration is shown in Figure 2-16. A telescope directs the incoming radiant flux obtained along a scan line through a scan line corrector to 1) the visible and near-infrared primary focal plane, or 2) the middle-infrared and thermal-infrared cooled focal plane. The detectors for the visible and near-infrared bands (1–4) are four staggered linear arrays, each containing 16 silicon detectors. The two middle-infrared detectors are 16 indium-antimonide cells in a staggered linear array, and the thermal infrared detector is a four-element array of mercury-cadmium-telluride cells.

Landsat TM data have a ground-projected IFOV of 30 × 30 m for bands 1 through 5 and 7. The thermal infrared band 6 has a spatial resolution of 120 × 120 m. The TM spectral bands represent important departures from the bands found on the traditional MSS, also carried onboard Landsats 4 and 5. The original MSS bandwidths were selected based on their utility for vegetation inventories and geologic studies. Conversely, the TM bands were chosen after years of analysis for their value in water penetration, discrimination of vegetation type and vigor, plant and soil moisture measurement, differentiation of clouds, snow, and ice, and identification of hydrothermal alteration in certain rock types (Table 2-3). The refined bandwidths and improved spatial resolution of the Landsat TM versus the Landsat MSS and several other sensor systems (Landsat 7 and SPOTs 1 – 4) are shown graphically in Figure 2-17. Examples of individual bands of Landsat Thematic Mapper imagery of Charleston, SC, obtained in 1994 are provided in Figure 2-18.

The Landsat TM bands were selected to make maximum use of the dominant factors controlling leaf reflectance, such as leaf pigmentation, leaf and canopy structure, and moisture content, as demonstrated in Figure 2-19. Band 1 (blue) provides water-penetration capability. Vegetation absorbs much of the incident blue, green, and red radiant flux for photosynthetic

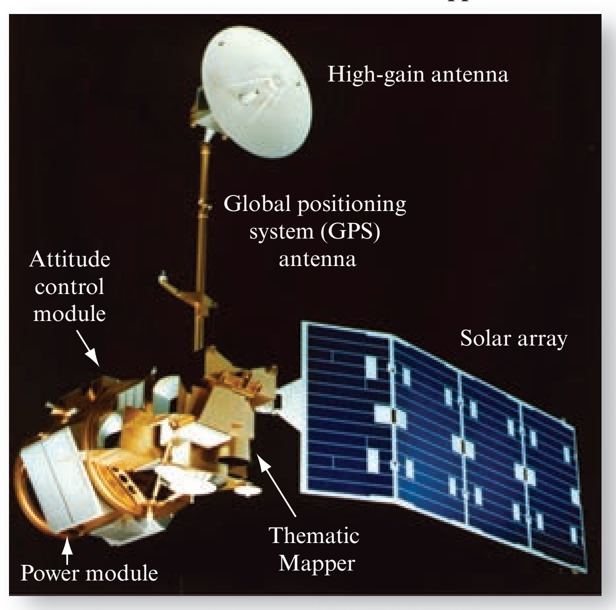


FIGURE 2–15 Landsat 4 and 5 platform and associated sensor and telecommunication systems.

purposes; therefore, vegetated areas appear dark in TM band 1 (blue), 2 (green), and 3 (red) images, as seen in the Charleston, SC, Landsat TM data (Figure 2-18). Vegetation reflects approximately half of the incident near-infrared radiant flux, causing it to appear bright in the band 4 (near-infrared) image. Bands 5 and 7 both provide more detail in the wetland because they are sensitive to soil and plant moisture conditions. The band 6 (thermal) image provides limited information of value.

The equatorial crossing time was 9:45 a.m. for Landsats 4 and 5 with an orbital inclination of 98.2°. The transition from a 919 km orbit to a 705 km orbit for Landsats 4 and 5 disrupted the continuity of Landsats 1, 2, and 3 MSS path and row designations in the Worldwide Reference System. Consequently, a separate WRS map is required to select images obtained by Landsats 4 and 5. The lower orbit (approximately the same as the space shuttle) also increased the amount of relief displacement introduced into the imagery obtained over mountainous terrain. The new orbit also caused the period between repetitive coverage to change from 18 to 16 days for both the MSS and TM data collected by Landsats 4 and 5.

There was a substantial improvement in the level of quantization from 6- to 8-bits per pixel (Table 2-2). This, in addition to a greater number of bands and a higher spatial resolution, increased the data rate from 15 to 85 Mb/s. Ground receiving stations were modified to process the increased data flow. Based on the

TABLE 2–3 Characteristics of the Landsat 4 and 5 Thematic Mapper (TM) spectral bands.

Band 1: 0.45 – 0.52 μm (blue). This band provides increased penetration of waterbodies, as well as supporting analyses of land-use, soil, and vegetation characteristics. The shorter wavelength cutoff is just below the peak transmittance of clear water, and the upper-wavelength cutoff is the limit of blue chlorophyll absorption for healthy green vegetation. Wavelengths < 0.45 μm are substantially influenced by atmospheric scattering and absorption.
Band 2: 0.52 – 0.60 μm (green). This band spans the region between the blue and red chlorophyll absorption bands and reacts to the green reflectance of healthy vegetation.
Band 3: 0.63 – 0.69 μm (red). This is the red chlorophyll absorption band of healthy green vegetation and is useful for vegetation discrimination. It is also useful for soil-boundary and geological-boundary delineations. This band may exhibit more contrast than bands 1 and 2 because of the reduced effect of atmospheric attenuation. The 0.69-μm cutoff is significant because it represents the beginning of a spectral region from 0.68 to 0.75 μm, where vegetation reflectance crossovers take place that can reduce the accuracy of vegetation investigations.
Band 4: 0.76 – 0.90 μm (near-infrared). For reasons discussed, the lower cutoff for this band was placed above 0.75 μm. This band is very responsive to the amount of vegetation biomass and/or leaf area present. It is useful for crop identification and emphasizes soil/crop and land/water contrasts.
Band 5: 1.55 – 1.75 μm (SWIR). This band is sensitive to the turgidity or amount of water in plants. Such information is useful in crop drought studies and in plant vigor investigations. This is one of the few bands that can be used to discriminate among clouds, snow, and ice.
Band 6: 10.4 – 12.5 μm (thermal infrared). This band measures the amount of infrared radiant energy emitted from surfaces. The apparent temperature is a function of the emissivities and the true (kinetic) temperature of the surface. It is useful for locating geothermal activity, thermal inertia mapping for geologic investigations, vegetation classification, vegetation stress analysis, and soil moisture studies. The band often captures unique information on differences in topographic aspect in mountainous areas.
Band 7: 2.08 – 2.35 μm (SWIR). This is an important band for the discrimination of geologic rock formations. It has been shown to be effective for identifying zones of hydrothermal alteration in rocks.

improvements in spectral, spatial, and radiometric resolution, Solomonson (1984) suggested that "it appears that the TM can be described as being twice as effective in providing information as the Landsat MSS. This is based on its ability to provide twice as many separable classes over a given area as the MSS, numerically provide two more independent vectors in the data or demonstrate through classical information theory that twice as much information exists in the TM data."

Efforts to move the Landsat program into the commercial sector began during the Carter administration in 1979 and resulted in 1984 legislation that charged the National Oceanic and Atmospheric Administration to transfer the program to the private sector. The Earth

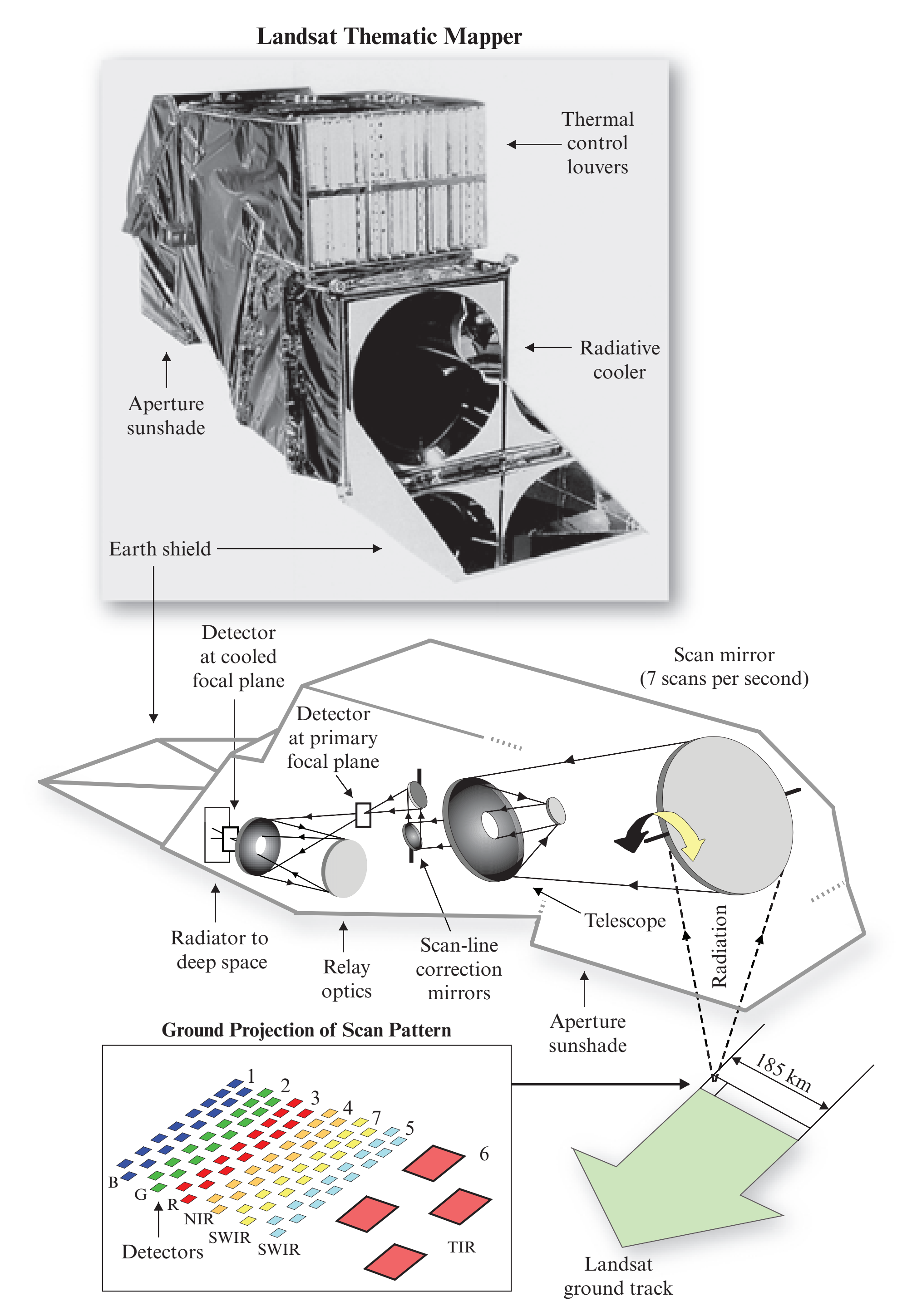


FIGURE 2–16 Major components of the Landsats 4 and 5 Thematic Mapper sensor system. The sensor is sensitive to the seven bands of the electromagnetic spectrum summarized in Table 2-2. Six of the seven bands have a spatial resolution of 30 × 30 m; the thermal infrared band has a spatial resolution of 120 × 120 m. The lower diagram depicts the sensor in its operational position.

Observing Satellite Company (EOSAT) took over operation in 1985 and was given the rights to market Landsat TM data.

Landsat 7 Enhanced Thematic Mapper Plus

On October 28, 1992, President Clinton signed the Land Remote Sensing Policy Act of 1992 (Public Law 102-555). This law authorized the procurement of Landsat 7 and called for its launch within 5 years of the launch of Landsat 6. In parallel actions, Congress funded Landsat 7 procurement and stipulated that data from publicly funded remote sensing satellite systems like Landsat must be sold to United States government agencies and their affiliated users at the cost

Spatial and Spectral Resolution of Selected Landsat, SPOT, Astrium Pleiades, and Digital Globe WorldView-2 Sensor Systems

FIGURE 2–17 Spatial and spectral resolution of the Landsat Multispectral Scanner (MSS), Landsats 4 and 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper Plus (ETM$^+$), Landsat 8 OLI, SPOTs 1, 2, and 3 High-Resolution Visible (HRV), SPOT 5 High-Resolution Visible Infrared (HRVIR) sensor, Pleiades HR-1, and DigitalGlobe WorldView-2 sensor systems. The SPOTs 4 and 5 *Vegetation* sensor characteristics are not shown (it consists of four 1.15 × 1.15 km bands).

of fulfilling user requests. Unfortunately, Landsat 6 did not achieve orbit on October 5, 1993.

With the passage of the Land Remote Sensing Policy Act of 1992, oversight of the Landsat program began to be shifted from the commercial sector back to the federal government. NASA was responsible for the design, development, launch, and on-orbit checkout of Landsat 7, and the installation and operation of the ground system. The U.S. Geological Survey (USGS) was responsible for data capture, processing, and distribution of the Landsat 7 data, mission management, and maintaining the Landsat 7 data archive.

Landsat 7 was launched on April 15, 1999, into a Sun-synchronous orbit (Figure 2-20). Landsat 7 was designed to work in harmony with NASA's EOS *Terra*

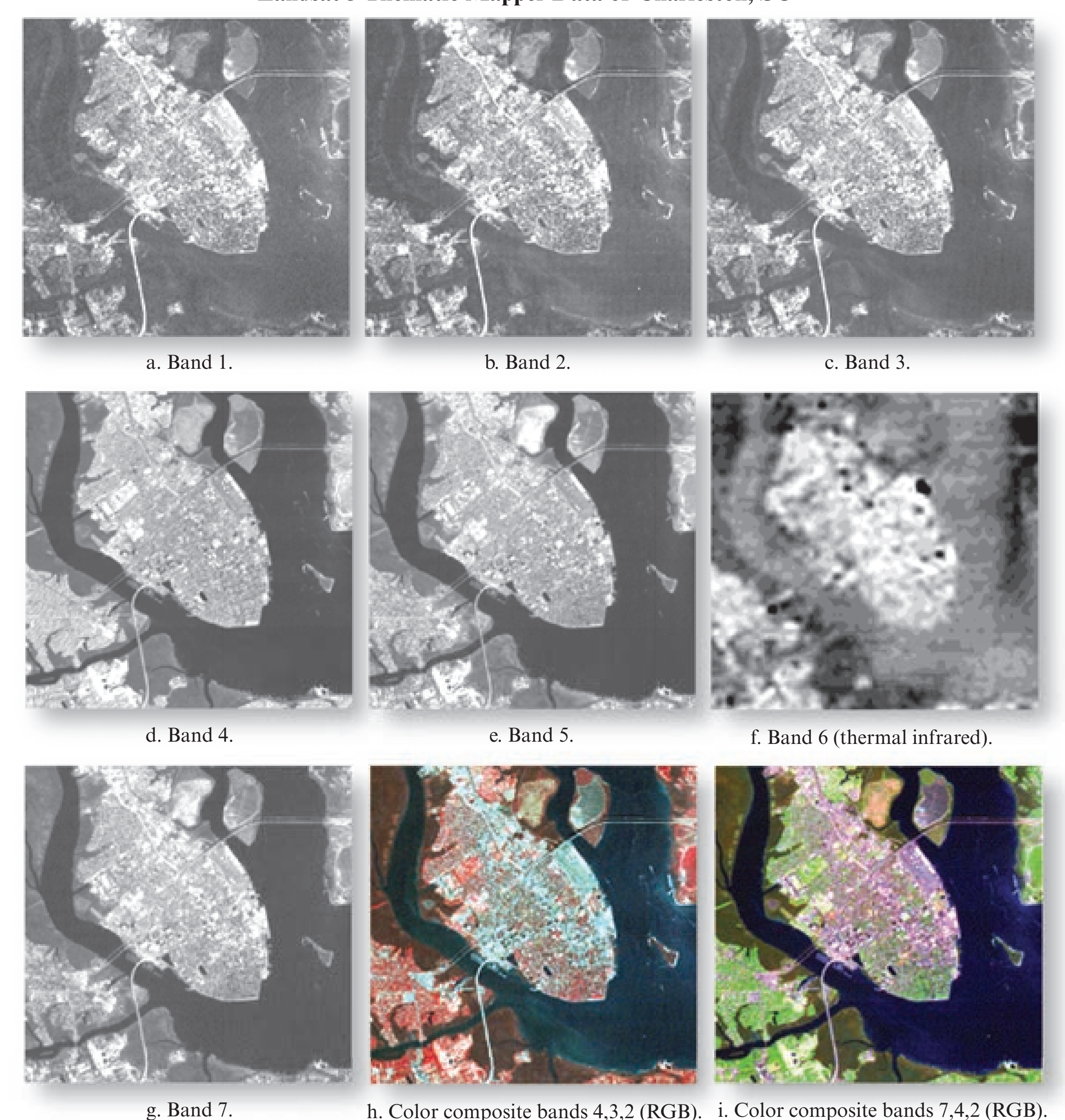


FIGURE 2–18 Landsat 5 Thematic Mapper data of Charleston, SC, obtained on February 3, 1994. Bands 1 through 5 and 7 are 30 × 30 m spatial resolution. Band 6 is 120 × 120 m (images courtesy of NASA).

satellite. It was also designed to achieve three main objectives:

- maintain data continuity by providing data that are consistent in terms of geometry, spatial resolution, calibration, coverage characteristics, and spectral characteristics with previous Landsat data;
- generate and periodically refresh a global archive of substantially cloud-free, sunlit landmass imagery; and
- continue to make Landsat-type data available to U.S. and international users at the cost of fulfilling user requests (COFUR) and to expand the use of such data for global-change research and commercial purposes.

Landsat 7 is a three-axis stabilized satellite platform carrying a single Nadir-pointing instrument, the ETM$^+$ (Figure 2-20). The ETM$^+$ instrument is a derivative of the Landsat 4 and 5 Thematic Mapper sensors. Therefore, it is possible to refer to Figure 2-16 for a review of its mirror and detector design. The ETM$^+$ was based on scanning technology despite the fact that linear array "pushbroom" technology had been commercially available since the launch of the French SPOT 1 satellite in 1986. Nevertheless, the ETM$^+$ instrument is

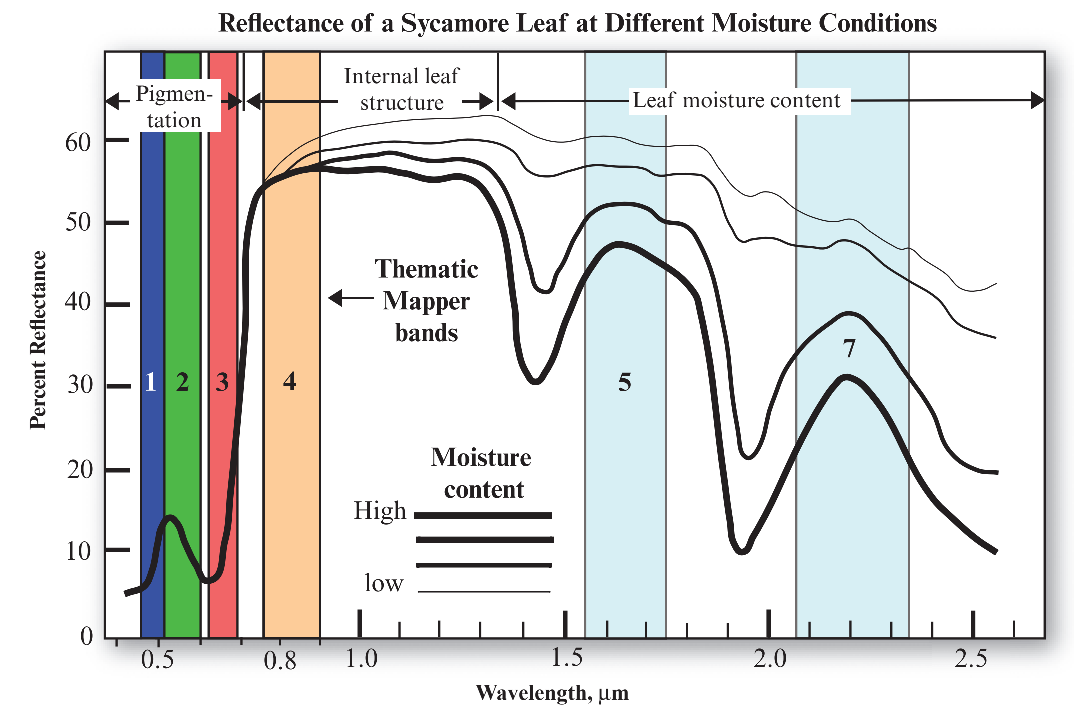


FIGURE 2–19 Progressive changes in percent reflectance for a sycamore leaf at varying oven-dry-weight moisture content. The dominant factors controlling leaf reflectance and the location of six of the Landsat TM bands are superimposed.

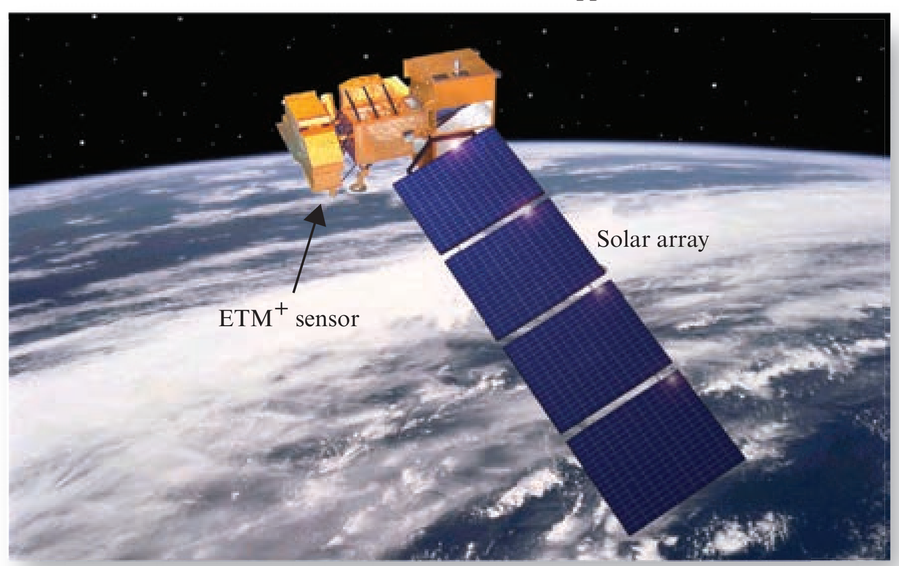


FIGURE 2–20 Artist's rendition of the Landsat 7 satellite with its Enhanced Thematic Mapper Plus (ETM^+) sensor system (courtesy of NASA).

an excellent sensor with several notable improvements over its predecessors, Landsat 4 and 5.

The characteristics of the Landsat 7 ETM^+ are summarized in Tables 2-3 and 2-4. The ETM^+ bands 1 through 5 and 7 are identical to those found on Landsats 4 and 5 and have the same 30 × 30 m spatial resolution. The thermal infrared band 6 has 60 × 60 m spatial resolution (instead of 120 × 120 m). Perhaps most notable was the new 15 × 15 m panchromatic band (0.52 – 0.90 μm). Landsat 7 ETM^+ individual band images of San Diego, CA, are shown in Figure 2-21. An ETM^+ color composite of San Diego is shown in Figure 2-22. Landsat 7 is in orbit 705 km above the Earth, collects data in a swath 185 km wide, and cannot view off-Nadir. Its revisit interval is 16 days. The ETM^+ records 150 megabits of data each second. Landsat 7 can transmit data to ground receiving sta-

TABLE 2–4 Landsat Enhanced Thematic Mapper Plus (ETM^+) compared with the Earth Observer (EO-1) sensors.

Landsat 7 Enhanced Thematic Mapper Plus (ETM^+)			EO-1 Advanced Land Imager (ALI)		
Band	**Spectral Resolution (μm)**	**Spatial Resolution (m) at Nadir**	**Band**	**Spectral Resolution (μm)**	**Spatial Resolution (m) at Nadir**
1 Blue	0.450 – 0.515	30 × 30	MS-1	0.433 – 0.453	30 × 30
2 Green	0.525 – 0.605	30 × 30	MS-1	0.450 – 0.510	30 × 30
3 Red	0.630 – 0.690	30 × 30	MS-2	0.525 – 0.605	30 × 30
4 Near-infrared	0.750 – 0.900	30 × 30	MS-3	0.630 – 0.690	30 × 30
5 SWIR	1.55 – 1.75	30 × 30	MS-4	0.775 – 0.805	30 × 30
6 Thermal IR	10.40 – 12.50	60 × 60	MS-4′	0.845 – 0.890	30 × 30
7 SWIR	2.08 – 2.35	30 × 30	MS-5′	1.20 – 1.30	30 × 30
8 Panchromatic	0.52 – 0.90	15 × 15	MS-5	1.55 – 1.75	30 × 30
			MS-7	2.08 – 2.35	30 × 30
			Panchromatic	0.480 – 0.69	10 × 10
			EO-1 Hyperion Hyperspectral Sensor 220 bands from 0.4 to 2.4 μm at 30 × 30 m		
			EO-1 LEISA Atmospheric Corrector (LAC) 256 bands from 0.9 to 1.6 μm at 250 × 250 m		
Sensor Technology	Scanning mirror spectrometer		Advanced Land Imager is a pushbroom radiometer. Hyperion is a pushbroom spectroradiometer. LAC uses area arrays.		
Swath Width	185 km		ALI = 37 km; Hyperion = 7.5 km; LAC = 185 km		
Data Rate	250 images per day @ 31,450 km^2		—		
Revisit	16 days		16 days		
Orbit and Inclination	705 km, Sun-synchronous Inclination = 98.2° Equatorial crossing 10:00 a.m. ±15 min.		705 km, Sun-synchronous Inclination = 98.2° Equatorial crossing = Landsat 7 + 1 min		
Launch	April 15, 1999		November 21, 2000		

tions at the EROS Data Center in Sioux Falls, SD, or to Fairbanks, AK. Landsat 7 international data may be acquired by retransmission using TDRS satellites or by international receiving stations.

The Landsat 7 ETM^+ has excellent radiometric calibration, which is accomplished using partial and full aperture solar calibration. Ground look calibration is performed by acquiring images of certain Earth landmass calibration targets. Biophysical and atmospheric characteristics of these targets are well instrumented on the ground.

At one time approximately 250 images were processed by the EROS Data Center each day. Unfortunately, the ETM^+ Scan Line Corrector (SLC) failed on May 31, 2003, resulting in imagery with significant geometric error. The SLC compensates for the forward motion of

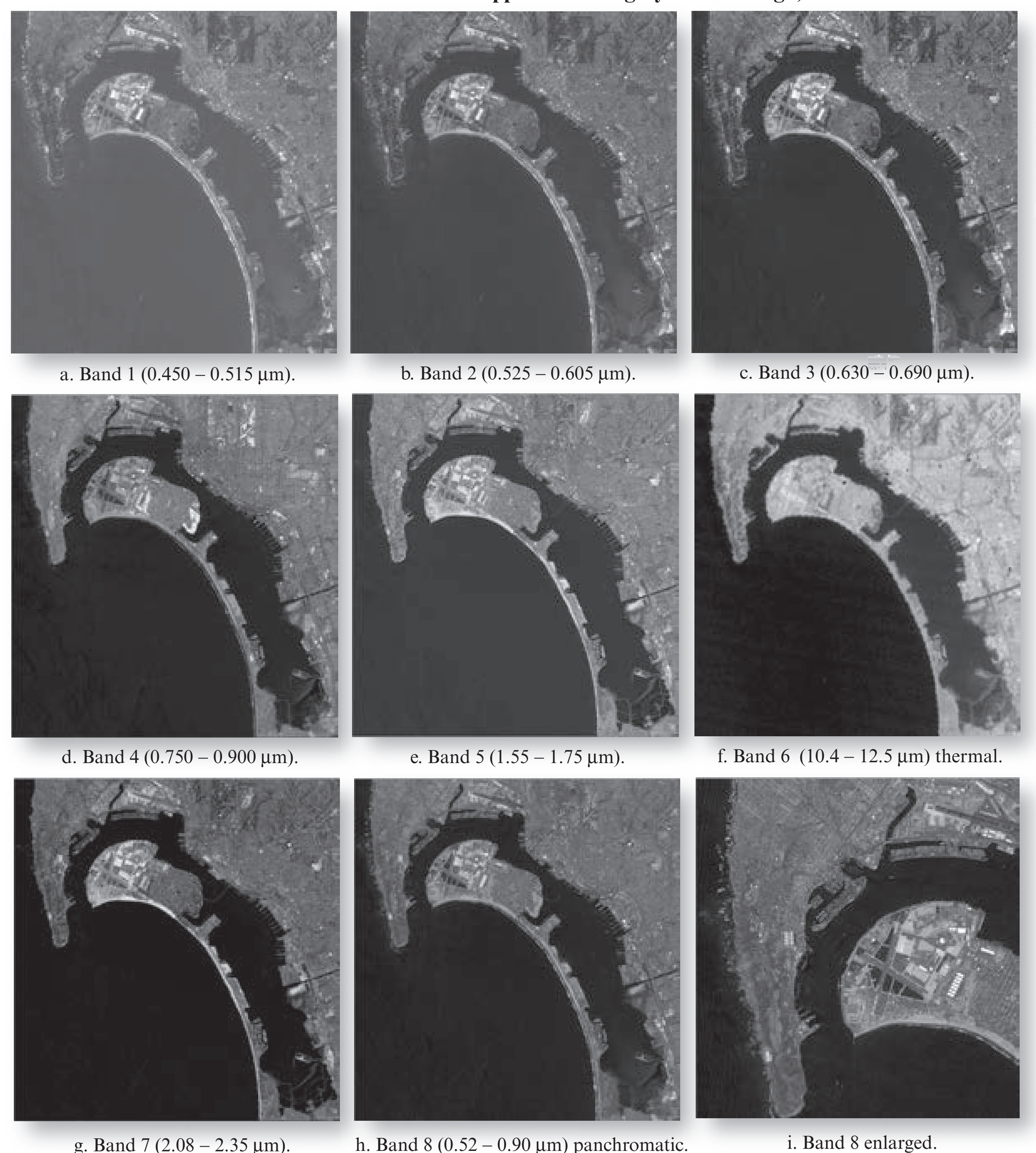


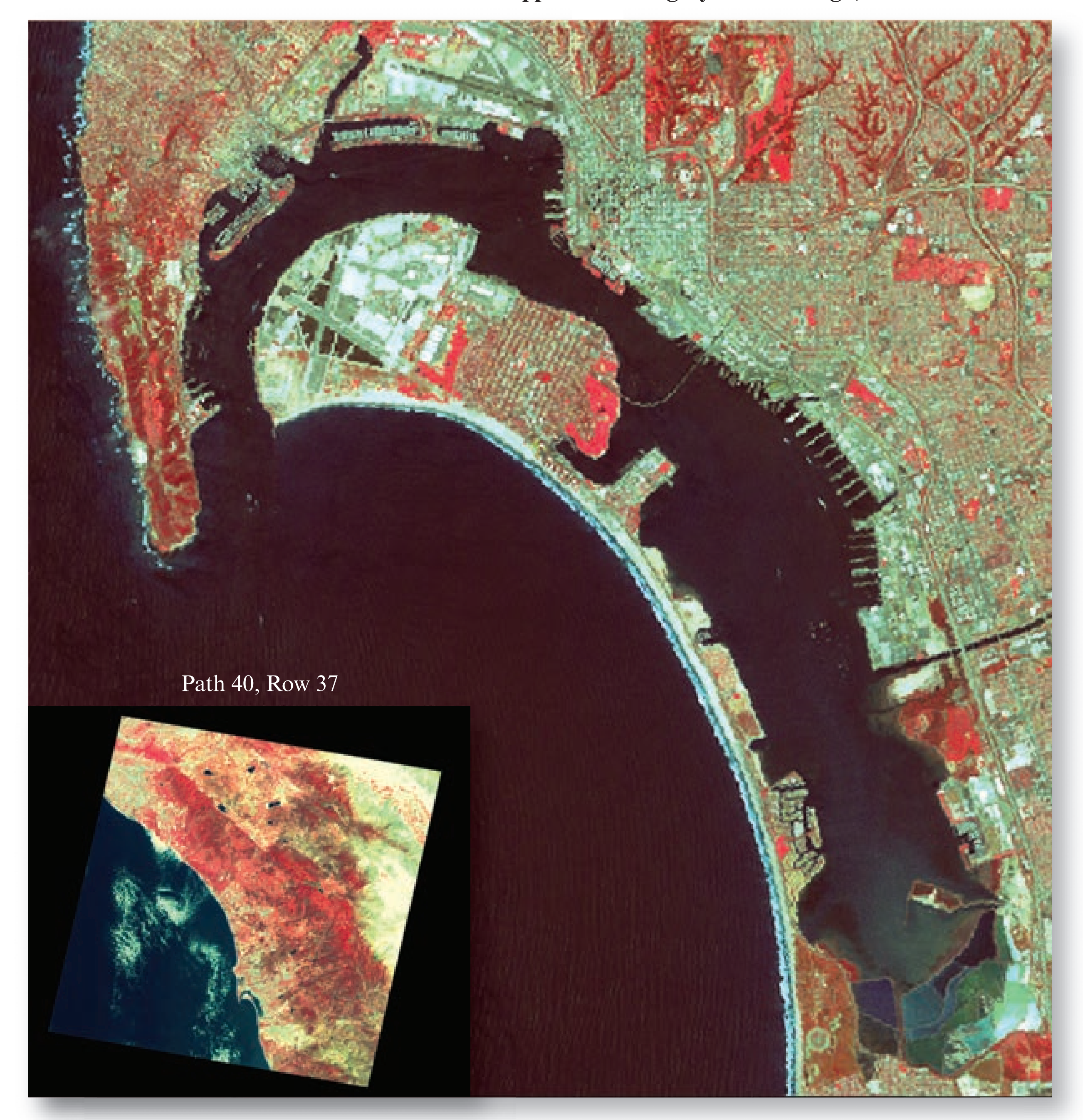
FIGURE 2–21 Landsat 7 ETM$^+$ imagery of San Diego, CA, obtained on April 24, 2000. Bands 1 through 5 and 7 are 30 × 30 m. Thermal infrared band 6 is 60 × 60 m. The panchromatic band 8 is 15 × 15 m (imagery courtesy of USGS and NASA).

the satellite. Efforts to recover the SLC were not successful. Portions of ETM$^+$ scenes obtained after this date in SLC-off mode are usable after special processing (USGS Landsat 7, 2004).

A USGS/NASA Landsat team refined ETM$^+$ gap-filling techniques that merge data from multiple ETM$^+$ acquisitions. They also developed modifications to the Landsat-7 acquisition scheme to acquire two or more clear scenes as near in time as possible to facilitate this gap-filling process. These merged images resolve some of the missing data problems (NASA Landsat 7, 2014).

Landsat 8 (LDCM–the Landsat Data Continuity Mission) launched on February 11, 2013, uses linear array technology and is discussed in a subsequent section.

Landsat 7 Enhanced Thematic Mapper Plus Imagery of San Diego, CA



Color-infrared color composite (RGB = Landsat ETM$^+$ bands 4, 3, 2).

FIGURE 2–22 Color composite of Landsat 7 Enhanced Thematic Mapper Plus imagery of San Diego, CA, obtained on April 24, 2000 (imagery courtesy of USGS and NASA).

NOAA Multispectral Scanner Sensors

NOAA operates two series of remote sensing satellites: the Geostationary Operational Environmental Satellites (GOES) and the Polar-orbiting Operational Environmental Satellites (POES). Both are currently based on multispectral scanner technology. The U.S. National Weather Service uses data from these sensors to forecast the weather. We often see GOES images of North and South America weather patterns on the daily news. The Advanced Very High Resolution Radiometer (AVHRR) was developed for meteorological purposes. However, global climate change research has focused attention on the use of AVHRR data to map vegetation and sea-surface characteristics.

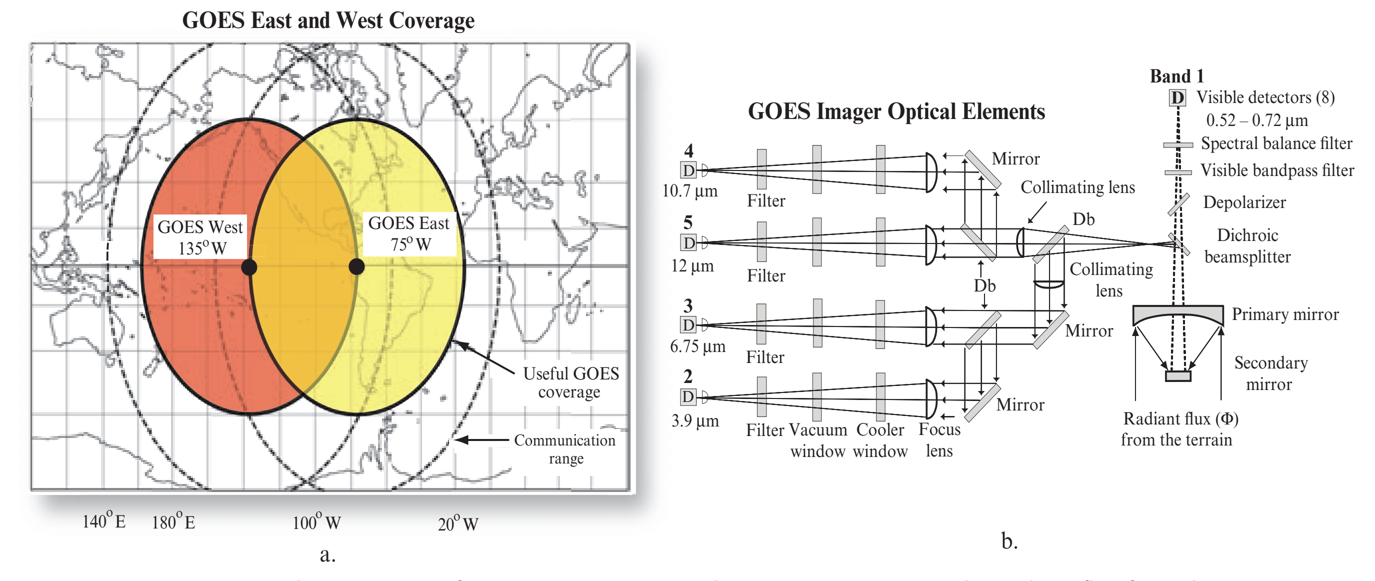


FIGURE 2–23 a) Geographic coverage of GOES East (75° W) and GOES West (135° W). b) Radiant flux from the terrain is reflected off a scanning mirror (not shown) onto the primary and secondary mirrors. A dichroic beamsplitter separates the visible light from the thermal infrared energy. Subsequent beamsplitters separate the thermal energy into specific bands.

Geostationary Operational Environmental Satellite (GOES)

The GOES system is operated by the National Environmental Satellite Data and Information Service (NESDIS) of NOAA. The system was developed by NESDIS in conjunction with NASA. GOES-N became operational on April 4, 2010, as GOES-13 East. Unfortunately, it failed on September 24, 2012, and GOES-O the backup satellite, was activated to become GOES-14 East. GOES P became operational on December 6, 2011, and became GOES-15 West (NOAA GOES, 2014).

GOES satellites are a mainstay of weather forecasting in the United States. They are the backbone of short-term forecasting or nowcasting. The real-time weather data gathered by GOES satellites, combined with data from Doppler radars and automated surface observing systems, greatly aid weather forecasters in providing warnings of thunderstorms, winter storms, flash floods, hurricanes, and other severe weather. These warnings help to save lives and preserve property

The GOES system consists of several observing subsystems:

- GOES Imager (provides multispectral image data),
- GOES Sounder (provides hourly 19-channel soundings), and
- a data-collection system (DCS) that relays data from *in situ* sites at or near the Earth's surface to other locations.

The GOES spacecraft is a three-axis (*x, y, z*) stabilized design capable of continuously pointing the optical line of sight of the imaging and sounding radiometers toward the Earth. GOES are placed in geostationary orbits approximately 35,790 km (22,240 statute miles) above the equator. The satellites remain at a stationary point above the equator and rotate at the same speed and direction as Earth. This enables the sensors to stare at a portion of the Earth from the geosynchronous orbit and thus more frequently obtain images of clouds, monitor the Earth's surface temperature and water vapor characteristics, and sound the Earth's atmosphere for its vertical thermal and water vapor structures.

The GOES East satellite is normally situated at 75°W longitude and GOES West is at 135°W longitude. The geographic coverage of GOES East and GOES West is summarized in Figure 2-23a. These sensors view most of the Earth from approximately 20°W to 165°E longitude. Poleward coverage is between approximately 77° N and S latitude. GOES East and West view the contiguous 48 states, South America, and major portions of the central and eastern Pacific Ocean and the central and western Atlantic Ocean. Pacific coverage includes the Hawaiian Islands and Gulf of Alaska, the latter known to weather forecasters as "the birthplace of North American weather systems."

GOES Imager The Imager is a five-channel multispectral scanner. The bandwidths and spatial resolution are summarized in Table 2-5. By means of a two-axis gimballed mirror in conjunction with a 31.1 cm (12.2 in) diameter Cassegrain telescope, the Imager's

TABLE 2–5 NOAA Geostationary Operational Environmental Satellite (GOES) Imager Instrument characteristics for GOES I-M (NOAA GOES, 2014).

Typical GOES Bands	Spectral Resolution (μm)	Spatial Resolution (km)	Band Utility
1 visible	0.55 – 0.75	1 × 1	Clouds, pollution, haze detection, and identification of severe storms
2 SWIR	3.80 – 4.00	4 × 4	Fog detection, discriminates between water, clouds, snow or ice clouds during daytime, detects fires and volcanoes, nighttime sea surface temperature (SST)
3 Moisture	6.50 – 7.00	8 × 8	Estimation of mid- and upper-level water vapor, detects advection, and tracks mid-level atmospheric motion
4 IR 1	10.2 – 11.2	4 × 4	Cloud-drift winds, severe storms, cloud-top heights, heavy rainfall
5 IR 2	11.5 – 12.5	4 × 4	Identification of low-level water vapor, SST, and dust and volcanic ash

multispectral channels can simultaneously sweep an 8 km (5 statute mile) north-to-south swath along an east-to-west/west-to-east path, at a rate of 20° (optical) per second. This translates into being able to scan a 3,000 by 3,000 km (1,864 by 1,864 miles) "box" centered over the United States in just 41 seconds. The actual scanning sequence takes place by sweeping in an east-west direction, stepping in the north-south direction, than sweeping back in a west-east direction, stepping north-south, sweeping east-west, and so on (NOAA GOES, 2014). The telescope concentrates both the visible and thermal radiant flux from the terrain onto a secondary mirror (Figure 2-23b). Dichroic beamsplitters separate the incoming scene radiance and focus it onto 22 detectors (8 visible and 14 thermal). The visible energy passes through the initial beamsplitter and is focused onto 8 silicon visible detector elements. Each of the 8 visible detectors has an IFOV of approximately 1 × 1 km at the satellite's suborbital point on the Earth.

All thermal infrared energy is deflected to the specialized detectors in the radiative cooler. The thermal infrared energy is further separated into the 3.9-, 6.75-, 10.7-, and 12- μm channels. Each of the four infrared channels has a separate set of detectors: four-element indium-antimonide (InSb) detectors for band 2; two-element mercury-cadmium-telluride (Hg:Cd:Te) detectors for band 3; and four-element mercury-cadmium-telluride (Hg:Cd:Te) detectors for both bands 4 and 5.

The GOES channels have 10-bit radiometric precision. The primary utility of the visible band 1 (1 × 1 km) is in the daytime monitoring of thunderstorms, frontal systems, and tropical cyclones. Band 2 (4 × 4 km) responds to both emitted terrestrial radiation and reflected solar radiation. It is useful for identifying fog and discriminating between water and ice clouds, and between snow and clouds, and for identifying large or very intense fires. It can be used at night to track low-level clouds and monitor near-surface wind circulation. Band 3 (8 × 8 km) responds to mid- and upper-level water vapor and clouds. It is useful for identifying the jet stream, upper-level wind fields, and thunderstorms. Energy recorded by band 4 (4 × 4 km) is not absorbed to any significant degree by atmospheric gases. It is ideal for measuring cloud-top heights, identifying cloud-top features, assessing the severity of some thunderstorms, and tracking clouds and frontal systems at night. Thermal band 5 (4 × 4 km) is similar to band 4 except that this wavelength region has a unique sensitivity to low-level water vapor. GOES-8 East visible, thermal infrared, and water vapor images of Hurricane Bonnie on August 25, 1998, are shown in Figure 2-24a–e. GOES-12 visible imagery of Hurricane Katrina collected on August 29, 2005, and draped over a MODIS color composite is shown in Figure 2-25a. GOES-12 visible imagery obtained on August 28, 2005, is shown in Figure 2-25b.

The Imager scans the continental United States every 15 minutes; scans most of the hemisphere from near the North Pole to approximately 20° S latitude every 26 minutes; and scans the entire hemisphere once every 3 hours in "routine" scheduling mode. Optionally, special imaging schedules are available, which allow data collection at more rapid time intervals over reduced geographic areas. GOES-14 is providing 1 minute imagery, which is giving scientists a sense of the type of imagery that GOES-R will provide after its launch in 2015. As amazing as 1-minute imagery appears, GOES-R will have the capability to provide 30-second imagery, which is 60× more frequent than GOES provides routine imagery (NOAA GOES, 2014).

GOES Sounder The GOES Sounder uses one visible and 18 infrared sounding channels to record data in a north-to-south swath across an east-to-west path. The Sounder and Imager both provide full Earth imagery,

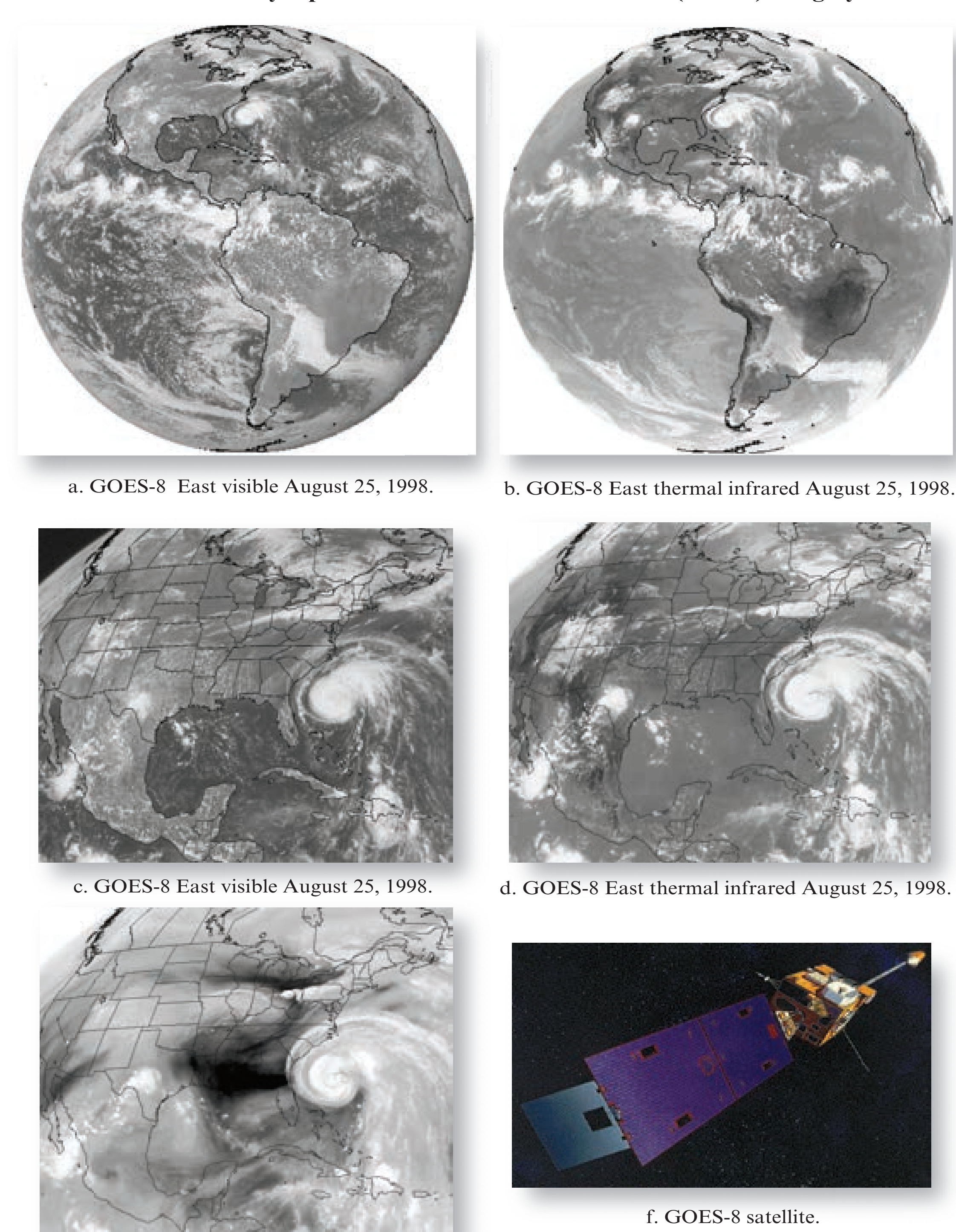
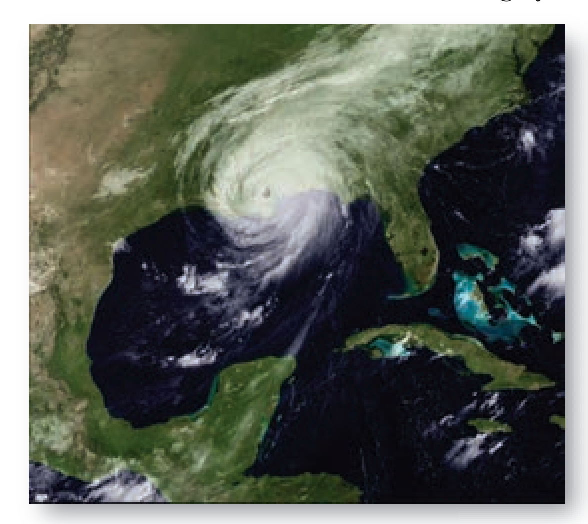


FIGURE 2–24 a–e) Examples of GOES-8 imagery obtained on August 25, 1998. f) The GOES-8 satellite (images courtesy of NOAA).

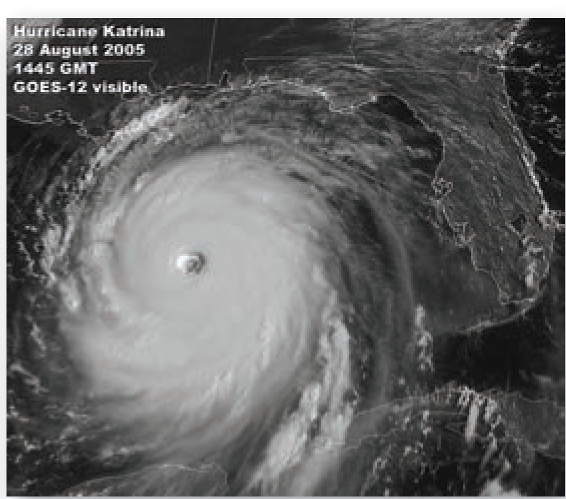
sector imagery, and area scans of local regions. The 19 bands yield the prime sounding products of vertical atmospheric temperature profiles, vertical moisture profiles, atmospheric layer mean temperature, layer mean moisture, total precipitable water, and the lifted index (a measure of stability). These products are used to augment data from the Imager to provide information on atmospheric temperature and moisture profiles, surface and cloud-top temperatures, and the distribution of atmospheric ozone.

GOES Imagery of Hurricane Katrina





a. GOES-12 visible band image draped over MODIS color composite on August 29, 2005.



b. GOES-12 visible band imagery recorded on August 28, 2005.

FIGURE 2–25 a) GOES-12 band 1 visible image of Hurricane Katrina on August 29, 2005, draped over a MODIS color composite (image courtesy of GOES Project NASA Goddard Space Flight Center). b) GOES-12 band 1 visible image of Hurricane Katrina on August 28, 2005 (image courtesy of NOAA).

Advanced Very High Resolution Radiometer

The Satellite Services Branch of the National Climatic Data Center, under the auspices of NESDIS, has established a digital archive of data collected from the NOAA Polar-orbiting Operational Environmental Satellites (POES). This series of satellites commenced with TIROS-N (launched in October, 1978) and continued with NOAA-A (launched in March 1983, and renamed NOAA-8) to the current NOAA-19 launched in 2009. These Sun-synchronous polar-orbiting satellites carry the *Advanced Very High Resolution Radiometer* (*AVHRR*). Substantial progress has been made in using AVHRR data for land-cover characterization and the mapping of daytime and nighttime clouds, snow, ice, and surface temperature. Unlike the Landsat TM and Landsat 7 ETM^+ sensor systems with nadir revisit cycles of 16 days, the AVHRR sensors acquire images of the entire Earth two times each day. This high frequency of coverage enhances the likelihood that cloud-free observations can be obtained for specific temporal windows and makes it possible to monitor change in land-cover conditions over short periods, such as a growing season. Moreover, the moderate resolution (1.1 × 1.1 km) of the AVHRR data makes it possible to collect, store, and process global datasets. For these reasons, NASA and NOAA initiated the AVHRR Pathfinder Program to create universally available global long-term remotely sensed datasets that can be used to study global climate change.

The AVHRR satellites orbit at approximately 833 km above Earth at an inclination of 98.9° and continuously record data in a swath 2,700 km wide at 1.1 × 1.1 km spatial resolution at nadir. Normally, two NOAA series satellites are operational at one time (one odd, one even). The odd-numbered satellite typically crosses the equator at approximately 2:30 p.m. and 2:30 a.m., and the even-numbered satellite crosses the equator at 7:30 p.m. and 7:30 a.m. local time. Each satellite orbits Earth 14.1 times daily (every 102 min) and acquires complete global coverage every 24 hours.

The AVHRR is a cross-track scanning system. The scanning rate of the AVHRR is 360 scans per minute. A total of 2,048 samples (pixels) are obtained per channel per Earth scan, which spans an angle of ±55.4° off-nadir. The IFOV of each band is approximately 1.4 milliradians leading to a resolution at the satellite subpoint of 1.1 × 1.1 km. The band characteristics of the various AVHRR satellites are summarized in Table 2-6. The sensitivity of five of the NOAA-19 bands is shown in Figure 2-26.

Full-resolution AVHRR data obtained at 1.1 × 1.1 km are called *local area coverage* (*LAC*) data. They may be resampled to 1.1 × 4 km *global area coverage* (*GAC*) data. The GAC data contain only one out of three original AVHRR lines and the data volume and resolution are further reduced by starting with the third sample along the scan line, averaging the next four samples, and skipping the next sample. The sequence of average four, skip one is continued to the end of the scan line. Some studies use GAC data while others use the full-resolution LAC data. Please be aware that the AVHRR effective spatial resolution becomes very coarse near the edges of the scan swath.

TABLE 2–6 NOAA Advanced Very High Resolution Radiometer sensor system characteristics (NOAA AVHRR, 2014; USGS AVHRR, 2014).

Band	NOAA-6, 8, 10, 12 Spectral Resolution (μm)[a]	NOAA-7, 9, 11, 13, 14 Spectral Resolution (μm)[a]	NOAA-15, 16, 17, 18, 19 AVHRR/3 Spectral Resolution (μm)[a]	Band Utility
1 Red	0.580 – 0.68	0.580 – 0.68	0.580 – 0.68	Daytime cloud, snow, ice, and vegetation mapping; used to compute NDVI
2 Near-IR	0.725 – 1.10	0.725 – 1.10	0.725 – 1.10	Land/water interface, snow, ice, and vegetation mapping; used to compute NDVI
3 SWIR	3.55 – 3.93	3.55 – 3.93	*3A*: 1.58 – 1.64 *3B*: 3.55 – 3.93	Snow and ice detection Nighttime cloud mapping and sea surface temperature
4 Thermal IR	10.50 – 11.50	10.30 – 11.30	10.30 – 11.30	Day/night cloud and surface-temperature mapping
5 Thermal IR	None	11.50 – 12.50	11.50 – 12.50	Cloud and surface temperature, day and night cloud mapping; removal of atmospheric water vapor path radiance
IFOV at nadir	1.1 × 1.1 km	1.1 × 1.1 km	1.1 × 1.1 km	
Swath width at nadir	2,700 km	2,700 km	2,700 km	

a. TIROS-N was launched on October 13, 1978; NOAA-6 on June 27, 1979; NOAA-7 on June 23, 1981; NOAA-8 on March 28, 1983; NOAA-9 on December 12, 1984; NOAA-10 on September 17, 1986; NOAA-11 on September 24, 1988; NOAA-12 on May 14, 1991; NOAA-13 on August 9, 1993; NOAA-14 on December 30, 1994; NOAA (K)-15 on May 13, 1998; NOAA (L)-16 on September 21, 2000; NOAA (M)-17 on June 24, 2002; NOAA (N)-18 on May 20, 2005; NOAA (M)-19 on February 6, 2009.

The AVHRR provides regional information on vegetation condition and sea-surface temperature. Band 1 is approximately equivalent to Landsat TM band 3. Vegetated land appears dark in this band of imagery due to chlorophyll absorption of red light. Band 2 is approximately equivalent to TM band 4. Vegetation reflects much of the near-infrared radiant flux while water absorbs much of the incident energy. The land–water interface is usually quite distinct. The three thermal bands provide information about Earth's surface and water temperature. For example, Figure 2-27 is a sea-surface temperature map derived from NOAA-16 AVHRR imagery obtained on October 16, 2003 (Gasparovic, 2003).

Scientists often compute a normalized difference vegetation index (NDVI) from the AVHRR data using the visible ($AVHRR_1$) and near-infrared ($AVHRR_2$) bands to map the condition of vegetation on a regional and national level. It is a simple transformation based on the following ratio:

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} = \frac{AVHRR_2 - AVHRR_1}{AVHRR_2 + AVHRR_1}. \quad (2.4)$$

The NDVI equation produces values in the range of –1.0 to 1.0, where increasing positive values indicate increasing green vegetation, and negative values indicate nonvegetated surfaces such as water, barren land, ice, and snow or clouds. To obtain the most precision, the NDVI is derived from calibrated, atmospherically corrected AVHRR channel 1 and 2 data in 16-bit precision, prior to geometric registration and sampling. The final NDVI results from –1 to 1 are normally scaled from 0 to 200. Vegetation indices are discussed in Chapter 8.

NDVI data obtained from multiple dates of AVHRR data can be composited to provide summary seasonal information. The *n*-day NDVI composite is produced by examining each NDVI value pixel by pixel for each observation during the compositing period to determine the maximum value. The retention of the highest NDVI value reduces the number of cloud-contaminated pixels.

The NDVI and other vegetation indexes (refer to Chapter 8) have been used extensively with AVHRR data to monitor natural vegetation and crop condition,

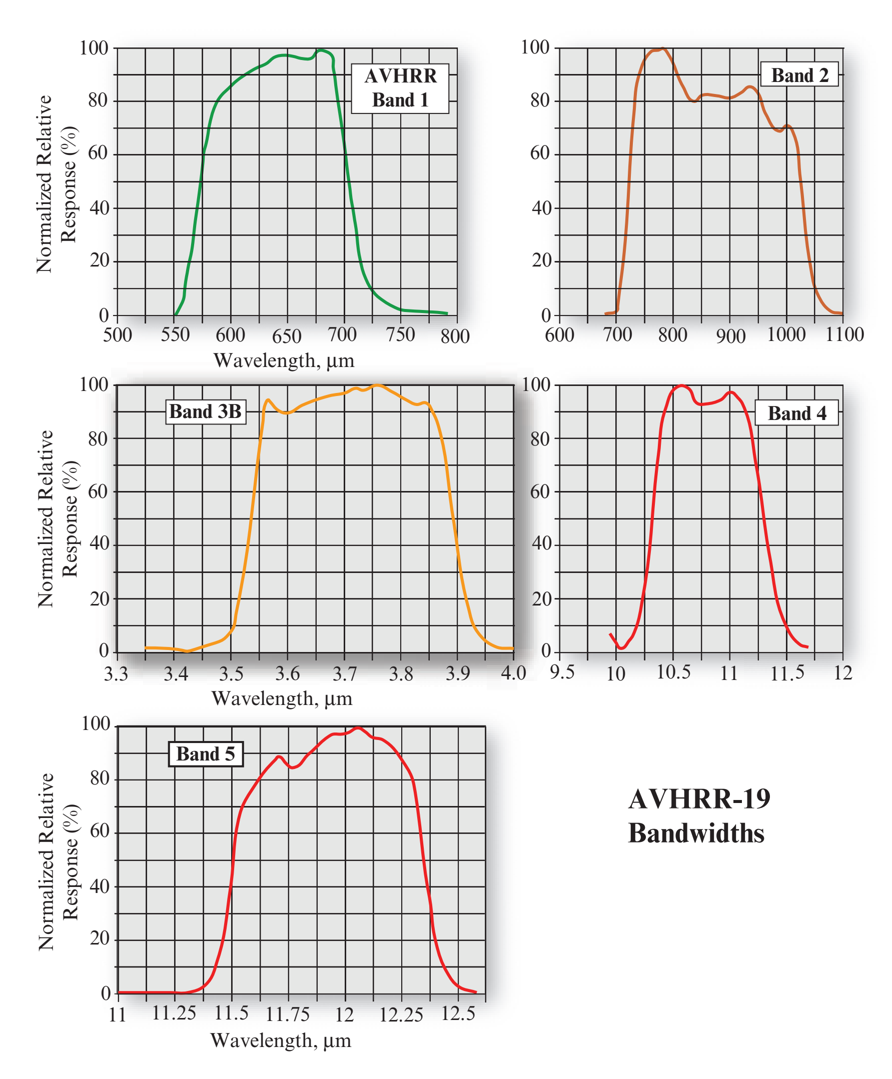


FIGURE 2–26 Characteristics of NOAA-19 AVHRR bandwidths.

identify deforestation in the tropics, and monitor areas undergoing desertification and drought. For example, the U.S. Geological Survey developed the Global Land Cover Characterization dataset based primarily on the unsupervised classification (refer to Chapter 9) of 1-km AVHRR 10-day NDVI composites. The AVHRR source imagery dates from April 1992 through March 1993. Ancillary data sources include digital elevation data, ecoregions interpretation, and country- or regional-level vegetation and land-cover maps (USGS Global Landcover, 2014).

NOAA Global Vegetation Index products based on AVHRR data are summarized as (NOAA GVI, 2014): first-generation (May 1982–April 1985); second-generation (April 1985–present), and third-generation new products (April 1985–present). NOAA and NASA are currently developing a follow-on to AVHRR and MODIS for the operational NPOESS constellation.

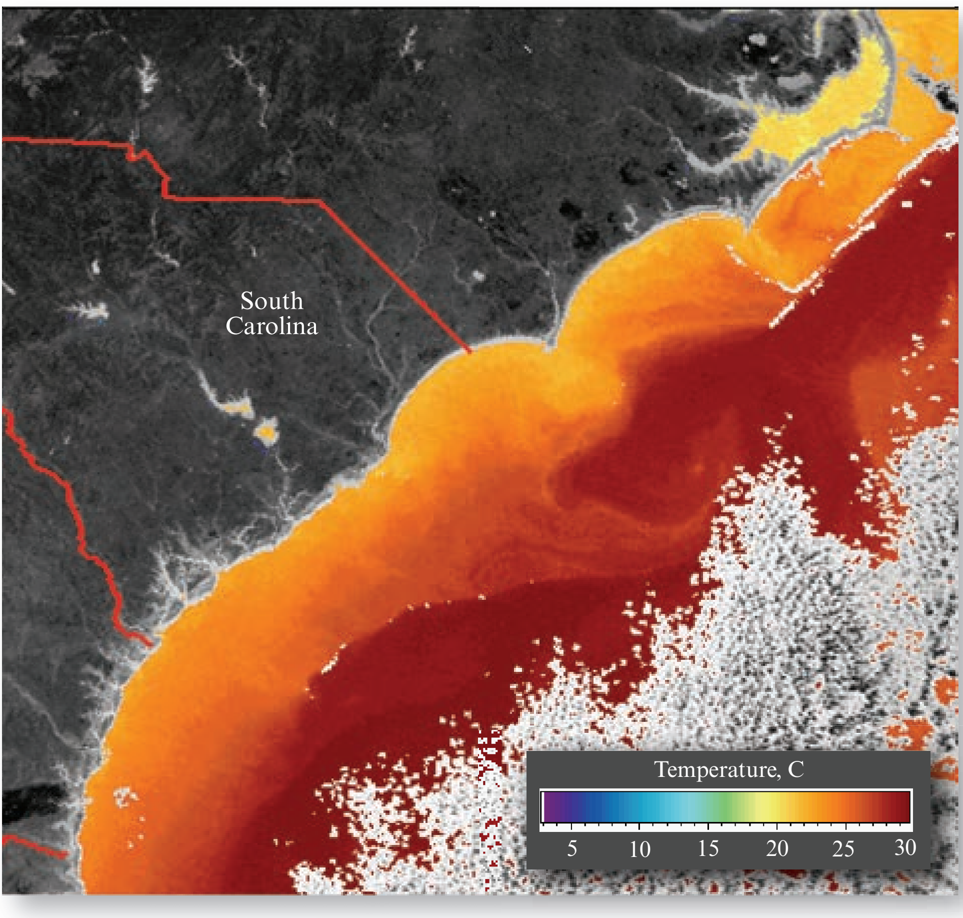
NOAA Suomi NPOESS Preparatory Project (NPP)

The Suomi NPOESS Preparatory Project (NPP) was launched October 28, 2011. It orbits the Earth about 14 times each day at 512 miles (824 km) above the Earth's surface. The Suomi NPP is the first in a new generation of satellites intended to replace the NOAA Polar Operational Environmental Satellites (POES) that were launched from 1997 to 2011. NPP observes the Earth's surface twice every 24-hour day, once in daylight and once at night.The NPP satellite is named after the late Verner E. Suomi of the University of Wisconsin.

The Suomi NPP satellite has five distinct imaging systems, including:

- The *Visible Infrared Imaging Radiometer Suite (VIIRS)* collects data in 22 channels between 0.4 and 12 μm using whiskbroom scanner technology and 12-bit quantization. VIIRS has a swath width of

NOAA-16 Advanced Very High Resolution Radiometer (AVHRR) Imagery



Sea-surface temperature (SST) map derived from NOAA-16 AVHRR band 4 (10.3 - 11.3 μm) imagery obtained on October 16, 2003.

FIGURE 2–27 Sea-surface temperature (SST) map derived from NOAA-16 AVHRR thermal infrared imagery (courtesy of NOAA and the Ocean Remote Sensing Program at Johns Hopkins University; Gasparovic, 2003).

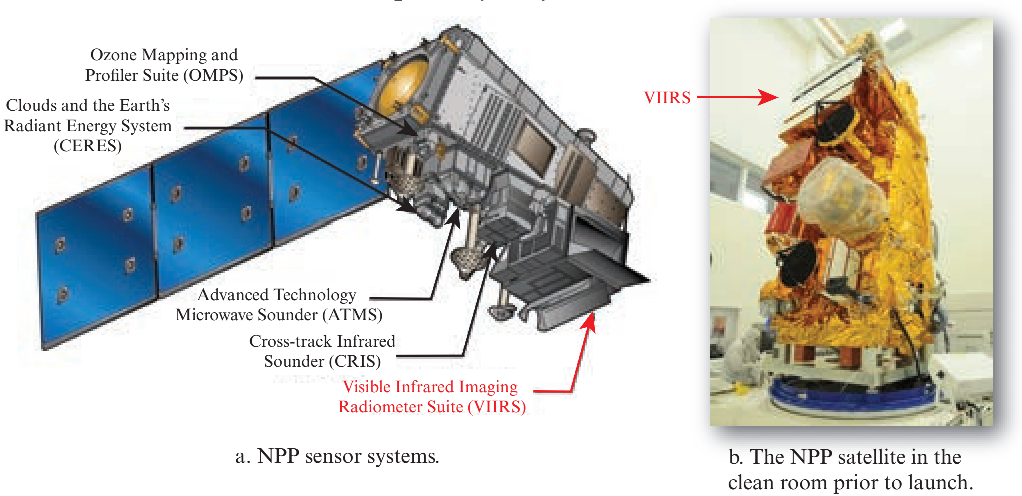
3,000 km. VIIRS has numerous bands that are very similar to those found on the Moderate Resolution Imaging Spectroradiometer (MODIS) operating on two NASA satellites, *Terra* and *Aqua* (NASA NPP, 2011; NASA VIIRS, 2014). A comparison of the VIIRS and MODIS bands is found in Table 2-7. Characteristics of the VIIRS platform are shown in Figure 2-28ab. The data are used to observe active fires, vegetation, ocean color, the urban heat-island, sea-surface temperature, and other surface features. A *Blue Marble* composite image of the Earth created from VIIRS imagery collected on January 4, 2012, is shown in Figure 2-28c.

- Advanced Technology Microwave Sounder (ATMS), a microwave radiometer which will help create global moisture and temperature models;
- Cross-Track Infrared Sounder (CrIS), a Michelson interferometer to monitor moisture and pressure;
- Ozone Mapping and Profiler Suite (OMPS), a group of imaging spectrometers to measure ozone levels, especially near the poles;
- Clouds and the Earth's Radiant Energy System (CERES), a radiometer to detect thermal radiation, including reflected solar radiation and thermal radiation emitted by the Earth.

VIIRS in effect replaces three currently operating sensors: the Defense Meteorological Satellite Program (DMSP) Operational Line-scanning System (OLS), the NOAA Polar-orbiting Operational Environmental Satellite (POES), Advanced Very High Resolution Radiometer (AVHRR), and the NASA Earth Observing System (EOS Terra and Aqua) Moderate-resolution Imaging Spectroradiometer (MODIS).

SeaStar Satellite and Sea-viewing Wide Field-of-view Sensor (SeaWiFS)

The oceans cover more than two-thirds of the Earth's surface and play an important role in the global climate system. *SeaWiFS* was an advanced scanning system designed specifically for ocean monitoring. The *SeaStar* satellite (*OrbView-2*) carried the SeaWiFS into



a. NPP sensor systems.

b. The NPP satellite in the clean room prior to launch.



c. NASA composite of multiple VIIRS images collected on January 4, 2012 into a "Blue Marble".

FIGURE 2–28 a) NPOESS Preparatory Project (NPP) satellite sensor systems. b) The NPP satellite in a clean room prior to launch. c) Blue Marble rendition of multiple NPP VIIRS images obtained on January 4, 2012 (NASA/NOAA/GSFC/Suomi NPP/VIIRS/Norman Kuring).

orbit using a Pegasus rocket on August 1, 1997 (NASA/Orbimage SeaWiFS, 2014). The final orbit was 705 km above the Earth. The equatorial crossing time was 12 p.m.

SeaWiFS

SeaWiFS built on all that was learned about ocean color remote sensing using the Nimbus-7 satellite Coastal Zone Color Scanner (CZCS) launched in 1978. CZCS ceased operation in 1986. The SeaWiFS instrument consisted of an optical scanner with a 58.3° total field of view. Incoming scene radiation was collected by a telescope and reflected onto the rotating half-angle mirror. The radiation was then relayed to dichroic beamsplitters that separate the radiation into eight wavelength intervals (Table 2-8). SeaWiFS had a spatial resolution of 1.13 × 1.13 km (at nadir) over a swath of 2,800 km. It had a revisit time of 1 day.

SeaWiFS recorded energy in eight spectral bands with very narrow wavelength ranges (Table 2-8) tailored for the detection and monitoring of very specific ocean phenomena, including ocean primary production and phytoplankton processes, ocean influences on climate processes (heat storage and aerosol formation), and the

TABLE 2–7 Characteristics of the NPOESS Preparatory Project (NPP) Visible Infrared Imaging Radiometer Suite (VIIRS) bands and their relationship with MODIS bands (Gleason et al., 2010).

VIIRS Band	Spectral Resolution (μm)	Spatial Resolution (m)	MODIS Bands	Spectral Resolution (μm)	Spatial Resolution (km)
DNB	0.5 – 0.9				
M1	0.402 – 0.422	750	8	0.405 – 0.420	1,000
M2	0.436 – 0.454	750	9	0.438 – 0.448	1,000
M3	0.478 – 0.498	750	3 or 10	0.459 – 0.479 0.483 – 0.493	500 1,000
M4	0.545 – 0.565	750	4 or 12	0.545 – 0.565 0.546 – 0.556	500 1,000
I1	0.600 – 0.680	375	1	0.620 – 0.670	250
M5	0.662 – 0.682	750	13 or 14	0.662 – 0.672 0.673 – 0.683	1,000 1,000
M6	0.739 – 0.754	750	15	0.743 – 0.753	1,000
I2	0.846 – 0.885	375	2	0.841 – 0.876	250
M7	0.846 – 0.885	750	16 or 2	0.862 – 0.877 0.841 – 0.876	1,000 250
M8	1.230 – 1.250	750	5	Same	500
M9	1.371 – 1.386	750	26	1.360 – 1.390	1,000
I3	1.580 – 1.640	375	6	1.628 – 1.652	500
M10	1.580 – 1.640	750	6	1.628 – 1.652	500
M11	2.225 – 2.275	750	7	2.105 – 2.115	500
I4	3.550 – 3.930	375	20	3.660 – 3.840	1,000
M12	3.550 – 3.930	750	20	same	1,000
M13	3.973 – 4.128	750	21 or 22	3.929 – 3.989 3.929 – 3.989	1,000 1,000
M14	8.400 – 8.700	750	29	same	1,000
M15	10.263 – 11.263	750	31	10.780 – 11.280	1,000
I5	10.500 – 12.400	375	31 or 32	10.780 – 11.280 11.770 – 12.270	1,000
M16	11.538 – 12.488	750	32	11.770 – 12.270	1,000

cycles of carbon, sulfur, and nitrogen. In particular, SeaWiFS had specially designed bands centered at 412 nm (to identify yellow substances through their blue wavelength absorption), at 490 nm (to increase sensitivity to chlorophyll concentration), and in the 765 and

TABLE 2–8 Characteristics of the Sea-viewing Wide Field-of-view Sensor.

SeaWiFS Band	Band Center (nm)	Band Width (nm)	Band Utility
1	412	402 – 422	Identify yellow substances
2	443	433 – 453	Chlorophyll concentration
3	490	480 – 500	Increased sensitivity to chlorophyll concentration
4	510	500 – 520	Chlorophyll concentration
5	555	545 – 565	Gelbstoffe (yellow substance)
6	670	660 – 680	Chlorophyll concentration
7	765	745 – 785	Surface vegetation, land–water interface, atmospheric correction
8	865	845 – 885	Surface vegetation, land–water interface, atmospheric correction

865 nm near-infrared (to assist in the removal of atmospheric attenuation).

SeaWiFS observations helped scientists understand the dynamics of ocean and coastal currents, the physics of mixing, and the relationships between ocean physics and large-scale patterns of productivity. The data filled the gaps in ocean biological observations between those of the test-bed CZCS and MODIS. SeaWiFS ceased operation on February 15, 2011. Information about obtaining historical SeaWiFS data can be found at http://oceancolor.gsfc.nasa.gov/SeaWiFS/.



Multispectral Imaging Using Linear Arrays

Linear array sensor systems use diodes or charge-coupled-devices to record the reflected or emitted radiance from the terrain. Linear array sensors are often called "pushbroom" sensors because, like a single line of bristles in a broom, the linear array stares constantly at the ground while the aerial or satellite platform moves forward (Figures 2-9d and 2-29a). The result is a more accurate measurement of the reflected radiant flux because 1) there is no moving mirror, and 2) the linear array detectors are able to dwell longer on a specific portion of the terrain.

NASA Earth Observing-1 (EO-1) Advanced Land Imager (ALI)

The NASA Earth Observing-1 (EO-1) satellite was launched on November 21, 2000, as part of a one-year technology validation/demonstration mission. The mission operated three advanced technology verification land imaging instruments. They were the first Earth-observing instruments to be flown under NASA's New Millennium Program. The three instruments are the Advanced Land Imager (ALI), the Hyperion hyperspectral imager, and the Linear Etalon Imaging Spectrometer Array (LEISA) Atmospheric Corrector (LAC). These instruments were supposed to enable future Landsat and Earth observing missions to more accurately classify and map land utilization globally. EO-1 was launched on November 21, 2001, into a polar orbit with an equatorial crossing time of 10:03 a.m. (descending node), an altitude of 705 km, an inclination of 98.2°, and an orbital period of 98 minutes. An agreement was reached between NASA and the USGS to allow continuation of the EO-1 Program as an Extended Mission. The EO-1 Extended Mission is chartered to collect and distribute ALI multispectral and Hyperion hyperspectral products in response to Data Acquisition Requests (DARs). Under the Extended Mission provisions, image data acquired by EO-1 are archived and distributed by the USGS Center for Earth Resources Observation and Science (EROS) and placed in the public domain (http://eo1.usgs.gov).

The Advanced Land Imager (ALI)

The ALI pushbroom instrument on EO-1 was used to validate and demonstrate technology for the Landsat Data Continuity Mission (LDCM). The ALI flies in formation with Landsat 7 ETM$^+$ (Digenis, 2005). The ALI has nine multispectral bands and a panchromatic band, three more than ETM$^+$, but does not have the thermal band (Mendenhall et al., 2012). It has increased sensitivity by a factor varying from four to ten

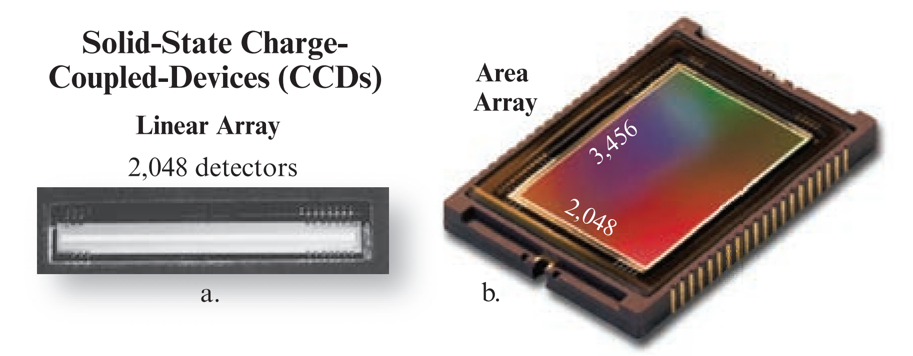


FIGURE 2–29 a) Enlarged view of a 2,048 element charge-coupled-device (CCD) linear array. b) Enlarged view of a 3,456 × 2,048 area array CCD.

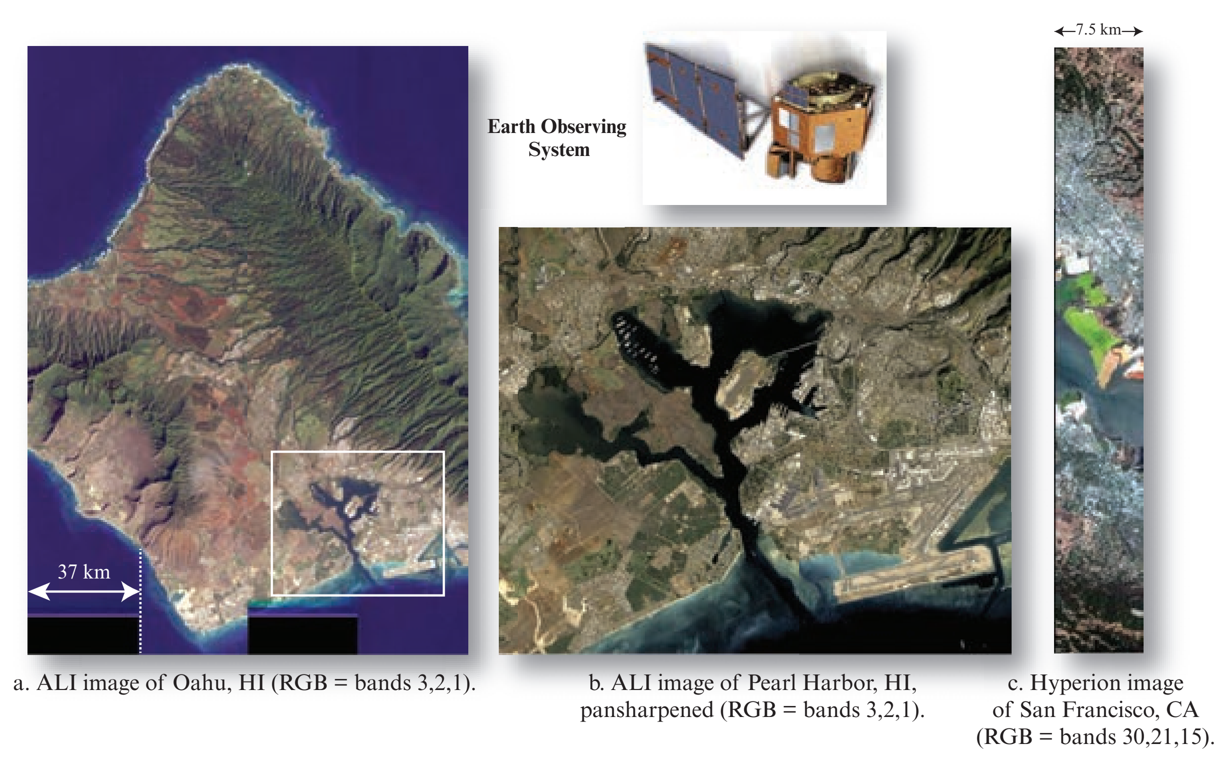


FIGURE 2–30 a) Mosaicked image of Oahu, HI, created using four 37 km swaths of EO-1 Advanced Land Imager (ALI) data. b) Enlargement of pan-sharpened ALI image of Pearl Harbor, HI. c) EO-1 Hyperion hyperspectral image of San Francisco, CA, with a swath width of 7.5 km (images courtesy of NASA Goddard Space Flight Center; Jensen et al., 2012).

depending upon the band. The spatial resolution of the multispectral bands is the same as that of ETM$^+$ (30 × 30 m) but the panchromatic band has improved spatial resolution (10 × 10 m versus 15 × 15 m). The wavelength coverage and ground sampling distance (GSD) are summarized in Table 2-9. Six of the nine multispectral bands are the same as those of the ETM$^+$ on Landsat 7, enabling direct comparison. The ALI has a swath width of 37 km. ALI images of Oahu, HI, are shown in Figure 2-30ab.

EO-1 specifications are summarized in Table 2-4. It contains a linear array Advanced Land Imager (ALI) with 10 bands from 0.4 to 2.35 μm at 30 × 30 m spatial resolution. The Hyperion hyperspectral sensor records data in 220 bands from 0.4 to 2.4 μm at 30 × 30 m spatial resolution. The Linear Etalon Imaging Spectrometer Array (LEISA) Atmospheric Corrector is a 256-band hyperspectral instrument sensitive to the region from 0.9 to 1.6 μm at 250 × 250 m spatial resolution. It is designed to correct for water-vapor variations in the

TABLE 2–9 Landsat 8 characteristics (Markham, 2011).

Landsat 8 Operational Land Imager (OLI)			Landsat 8 Thermal Infrared Sensor (TIRS)		
Band	**Spectral Resolution (μm)**	**Spatial Resolution (m) at Nadir**	**Band**	**Spectral Resolution (μm)**	**Spatial Resolution (m) at Nadir**
1 Ultra-blue for coastal/aerosol	0.433 – 0.453	30 × 30	10 Thermal	0.3 – 11.3	100 × 100
2 Blue	0.450 – 0.515	30 × 30	11 Thermal	11.5 – 12.5	100 × 100
3 Green	0.525 – 0.600	30 × 30	Sensor Technology	Pushbroom	
4 Red	0.630 – 0.680	30 × 30	Swath Width	185 km	
5 Near-infrared	0.845 – 0.885	30 × 30	Data Rate	400 WRS-2 scenes per day	
6 SWIR-1	1.56 – 1.66	30 × 30	Revisit	16 days	
7 SWIR-2	2.1 – 2.3	30 × 30	Orbit and Inclination	705 km, Sun-synchronous Inclination = 98.2° Equatorial crossing 10:00 a.m.	
8 Panchromatic	0.52 – 0.90	15 × 15	Quantization	12-bits (dramatic improvement over ETM^+)	
9 Cirrus	1.36 - 1.39	30 × 30	Signal to Noise	1–2 orders of magnitude improvement	

atmosphere. All three of the EO-1 land imaging instruments view all or sub-segments of the Landsat 7 swath.

NASA Landsat 8 (LDCM—Landsat Data Continuity Mission)

Landsat 8 (LDCM—the Landsat Data Continuity Mission) is a partnership between NASA and the USGS. Landsat 8 was launched on February 11, 2013. The goal of Landsat 8 is to: 1) collect and archive medium resolution (circa 15- to 30-m spatial resolution) multispectral image data affording seasonal coverage of the global landmasses for a period of no less than 5 years, 2) ensure that Landsat 8 data are sufficiently consistent with data from the earlier Landsat missions in terms of acquisition geometry, calibration, coverage characteristics, spectral characteristics, output product quality, and data availability to permit studies of land-cover and general land-use change over time, and 3) distribute Landsat 8 data products to the public via the Internet on a nondiscriminatory basis and at a price no greater than the incremental cost of fulfilling a user request (USGS, 2010).

Orbital Land Imager

Landsat 8 is planned as a 5-year mission but will include enough fuel for 10 years of operation. The multispectral remote sensing instrument is the Operational Land Imager (OLI). The OLI collects land-surface data with a spatial and spectral resolution consistent with historical Landsat TM and ETM^+ data previously discussed at 15 × 15 m (panchromatic) and 30 × 30 m (multispectral VNIR/SWIR) (Figure 2-17). It also has two additional spectral channels: an "ultra-blue" band for coastal and aerosol studies and a band for cirrus cloud detection. The OLI uses long detector arrays, with approximately 7,000 detectors per spectral band, aligned across its focal plane to view across the swath. The 15 × 15 m panchromatic band has more than 13,000 detectors. Silicon PIN (SiPIN) detectors that collect the data for the visible and near-infrared spectral bands (Bands 1 to 4 and 8) while mercury–cadmium–telluride (Hg:Cd:Te) detectors are used for the shortwave infrared bands (Bands 6, 7, and 9). This "pushbroom" design results in a more sensitive instrument providing improved land surface information with fewer moving parts (Irons et al., 2012).

The thermal infrared sensor (TIRS) collects data in two long wavelength bands that are co-registered with the OLI data. About 400 scenes per day will be collected and terrain-corrected to a geographic projection. OLI and TIRS data are distributed as a combined product. In 2007, a plan was put forth for a National

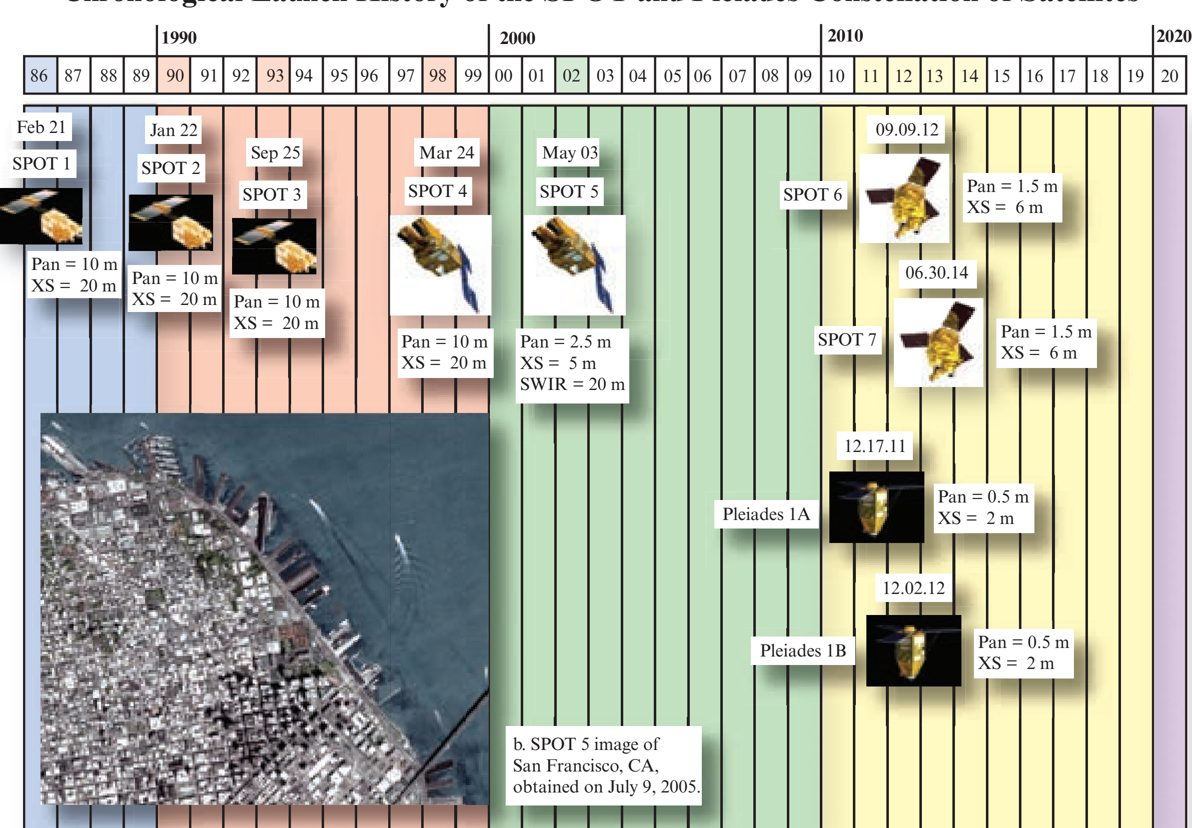


FIGURE 2–31 Chronological launch history of the SPOT satellites (updated from Jensen et al., 2012; SPOT 5 image courtesy of Airbus Defense and Space).

Land Imaging Program that includes calling for operational moderate-resolution land imaging into the future (USGS, 2010). Landsat 9 is in the preliminary planning stages for launch in approximately 2018.

SPOT Sensor Systems

The first SPOT satellite was launched on February 21, 1986 (Figure 2-31). It was developed by the French Centre National d'Etudes Spatiales (CNES) in cooperation with Belgium and Sweden. It had a spatial resolution of 10 × 10 m (panchromatic mode) and 20 × 20 m (multispectral mode) and provided several other innovations in remote sensor system design (Table 2-10). SPOT satellites 2 and 3 with identical payloads were launched on January 22, 1990, and September 25, 1993, respectively. SPOT 4 was launched on March 24, 1998, with a *Vegetation* sensor (1 × 1 km) and a shortwavelength infrared band. SPOT 5 was launched on May 3, 2002, with visible, near-infrared, and shortwave infrared (SWIR) bands at 10 × 10 m and a panchromatic band at 2.5 × 2.5 m. SPOT 6 was launched on September 9, 2012. SPOT was the first commercially-available satellite system based on linear array technology and the first pointable sensor.

Since 1986, SPOT Earth observation satellites have been a consistent, dependable source of high-resolution Earth resource information. While many countries have seen their Earth resource monitoring sensor systems come and go depending primarily upon politics, one could always count on SPOT Image, Inc. to provide quality imagery. Unfortunately, the imagery has always been relatively expensive, usually more than $2,000 per panchromatic or multispectral scene, although it has been reduced in recent years. The expense can be substantial if both panchromatic and multispectral imagery of the study area are required.

SPOT 1, 2, and 3

These satellites are all identical and consist of two parts: 1) the SPOT bus, which is a standard multipurpose platform, and 2) the sensor system instruments (Figure 2-32ab) consisting of two high-resolution visible (HRV) sensors and a package comprising two data recorders and a telemetry transmitter. The satellites operate in a Sun-synchronous, near-polar orbit (inclination of 98.7°) at an altitude of 822 km. The satellites pass overhead at the same solar time; the local clock time varies with latitude.

TABLE 2–10 SPOT 1, 2, and 3 High-Resolution Visible (HRV), SPOT 4 and 5 High-Resolution Visible and Infrared (HRVIR), and SPOT 4 and 5 *Vegetation* sensor system characteristics.

SPOT 1, 2, 3 HRV and 4 HRVIR			SPOT 5 HRVIR			SPOT 4 and 5 *Vegetation*		
Band	**Spectral Resolution (μm)**	**Spatial Resolution (m) at Nadir**	**Band**	**Spectral Resolution (μm)**	**Spatial Resolution (m) at Nadir**	**Band**	**Spectral Resolution (μm)**	**Spatial Resolution (km) at Nadir**
1	0.50 – 0.59	20 × 20	1	0.50 – 0.59	10 × 10	1	0.43 – 0.47	1.15 × 1.15
2	0.61 – 0.68	20 × 20	2	0.61 – 0.68	10 × 10	2	0.61 – 0.68	1.15 × 1.15
3	0.79 – 0.89	20 × 20	3	0.79 – 0.89	10 × 10	3	0.78 – 0.89	1.15 × 1.15
Pan Pan (4)	0.51 – 0.73 0.61 – 0.68	10 × 10 10 × 10	Pan	0.48 – 0.71	2.5 × 2.5			
SWIR (4)	1.58 – 1.75	20 × 20	SWIR	1.58 – 1.75	20 × 20	SWIR	1.58 – 1.75	1.15 × 1.15
Sensor	Linear array pushbroom		Linear array pushbroom			Linear array pushbroom		
Swath	60 km ± 50.5°		60 km ± 27°			2,250 km ± 50.5°		
Rate	25 Mb/s		50 Mb/s			50 Mb/s		
Revisit	26 days		26 days			1 day		
Orbit	822 km, Sun-synchronous Inclination = 98.7° Equatorial crossing 10:30 a.m.		822 km, Sun-synchronous Inclination = 98.7° Equatorial crossing 10:30 a.m.			822 km, Sun-synchronous Inclination = 98.7° Equatorial crossing 10:30 a.m.		

The HRV sensors operate in two modes in the visible and reflective-infrared portions of the spectrum. The first is a panchromatic mode corresponding to observation over a broad spectral band (similar to a typical black-and-white photograph). The second is a multispectral (color) mode corresponding to observation in three relatively narrow spectral bands (Table 2-10). Thus, the spectral resolution of SPOTs 1 through 3 is not as good as that of the Landsat Thematic Mapper. The ground spatial resolution, however, is 10 × 10 m for the panchromatic band and 20 × 20 m for the three multispectral bands when the instruments are viewing at nadir, directly below the satellite.

Reflected energy from the terrain enters the HRV via a plane mirror and is then projected onto two CCD arrays. Each CCD array consists of 6,000 detectors arranged linearly. An electron microscope view of some of the individual detectors in the linear array is shown in Figure 2-33ab. This linear array pushbroom sensor images a complete line of the ground scene in the cross-track direction in one look as the sensor system progresses downtrack (refer to Figure 2-5d). This capability breaks tradition with the Landsat MSS, Landsat TM, and Landsat 7 ETM^+ sensors because no mechanical scanning takes place. A linear array sensor is superior because there is no mirror that must scan back and forth to collect data (mirror-scan velocity is a serious issue) and this allows the detector to literally "stare" at the ground for a longer time, obtaining a more accurate record of the spectral radiant flux exiting the terrain. The SPOT satellites pioneered this linear array pushbroom technology in commercial Earth resource remote sensing as early as 1986.

When looking directly at the terrain beneath the sensor system, the two HRV instruments can be pointed to cover adjacent fields, each with a 60-km swath width (Figure 2-32c). In this configuration the total swath width is 117 km and the two fields overlap by 3 km. It is also possible to selectively point the mirror to off-Nadir viewing angles through commands from the ground station. In this configuration it is possible to observe any region of interest within a 950-km wide strip centered on the satellite ground track (i.e., the observed region may not be centered on the ground track) (Figure 2-34a). The width of the swath actually

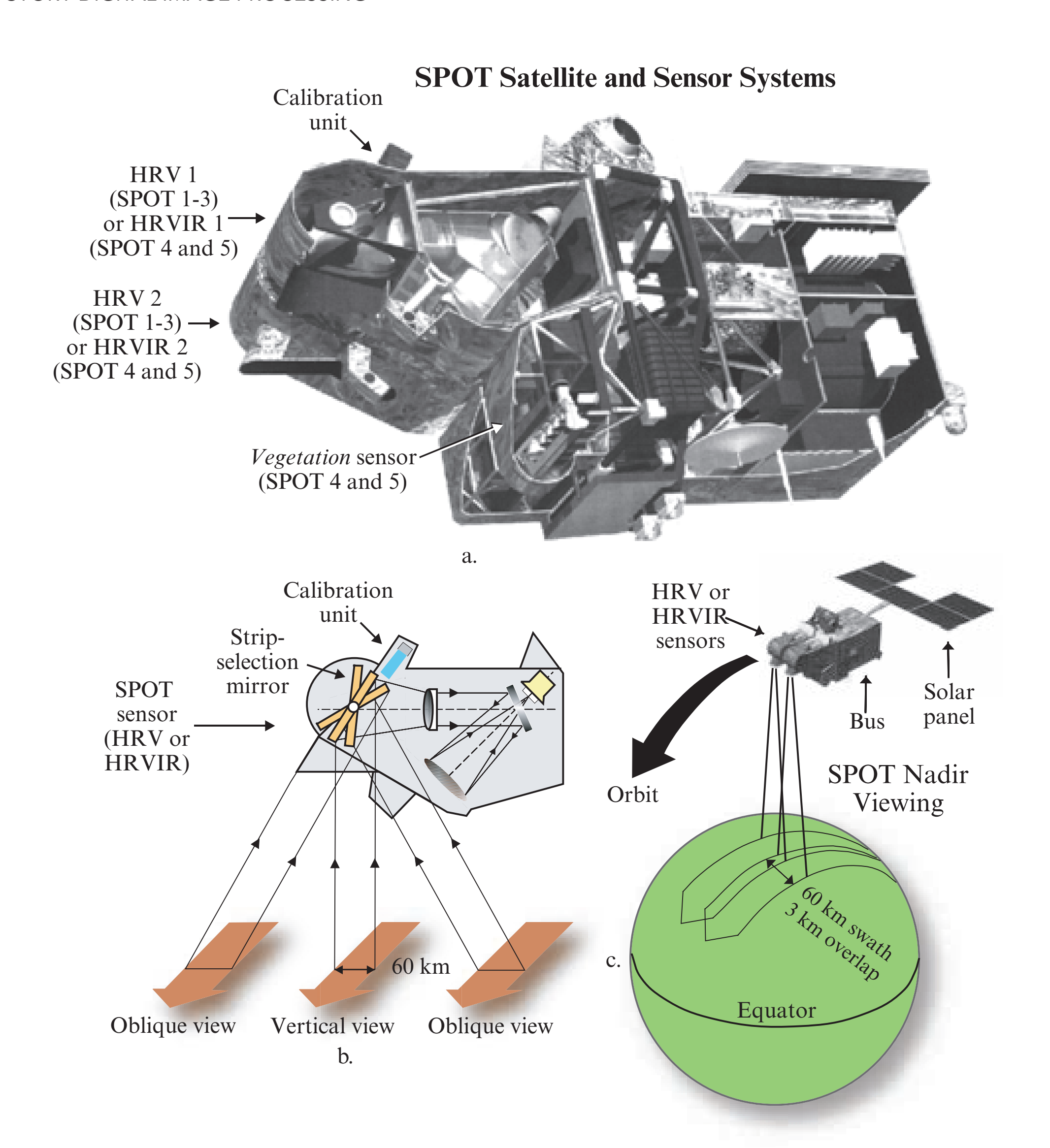


FIGURE 2–32 a–c) The SPOT satellites consist of the SPOT bus, which is a multipurpose platform, and the sensor system payload. Two identical high-resolution visible (HRV) sensors are found on SPOTs 1, 2, and 3 and two identical high-resolution visible infrared (HRVIR) sensors on SPOT 4. Radiant energy from the terrain enters the HRV or HRVIR via a plane mirror and is then projected onto two CCD arrays. Each CCD array consists of 6,000 detectors arranged linearly. This results in a spatial resolution of 10 × 10 or 20 × 20 m, depending on the mode in which the sensor is being used. The swath width at nadir is 60 km. The SPOT HRV and HRVIR sensors may also be pointed off-Nadir to collect data. SPOT 4 carries a *Vegetation* sensor with 1.15 × 1.15 km spatial resolution and 2,250-km swath width (adapted from SPOT Image, Inc.).

observed varies between 60 km for nadir viewing and 80 km for extreme off-nadir viewing.

If the HRV instruments were capable only of Nadir viewing, the revisit frequency for any given region of the world would be 26 days. This interval is often unacceptable for the observation of phenomena evolving on time scales ranging from several days to a few weeks, especially where cloud cover hinders the acquisition of usable data. During the 26-day period separating two successive SPOT satellite passes over a given point on Earth and taking into account the steering capability of the instruments, the point in question could be observed on seven different passes if it were on the equator and 11 times if at a latitude of 45° (Figure 2-34b). A given region can be revisited on dates separated alternatively by 1 to 4 (or occasionally 5) days.

The SPOT sensors can also acquire cross-track stereoscopic pairs of images of a geographic region (Figure 2-34c). Two observations can be made on successive days such that the two images are acquired at angles on either side of the vertical. In such cases, the ratio between the observation base (distance between the two satellite positions) and the height (satellite altitude) is approximately 0.75 at the equator and 0.50 at a latitude of 45°. SPOT data with these base-to-height ratios may be used for topographic mapping.

SPOT Linear Array Enlargements

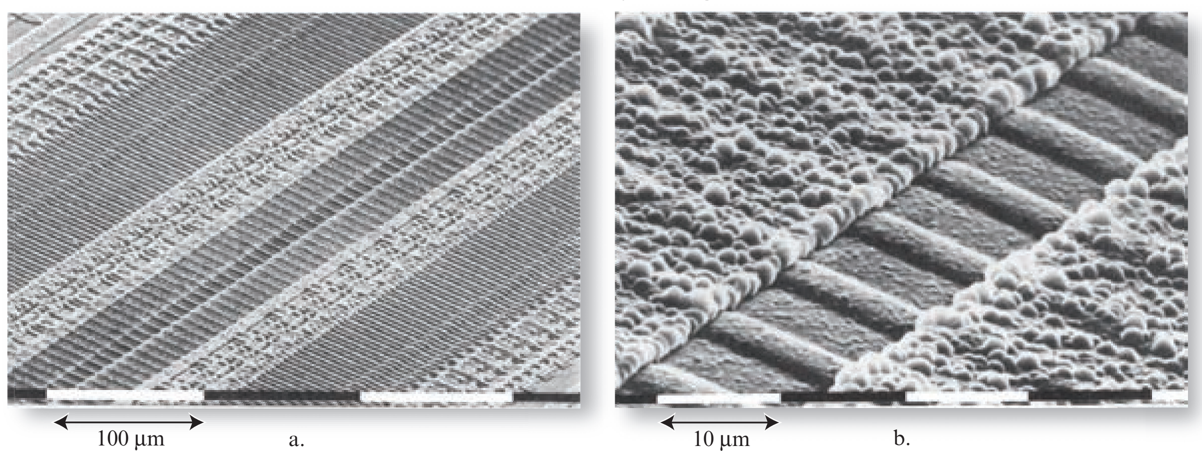


FIGURE 2–33 a) Scanning electron microscope images of the front surface of a CCD linear array like that used in the SPOT HRV sensor systems. Approximately 58 CCD detectors are visible, with rows of readout registers on both sides. b) Seven detectors of a CCD linear array are shown at higher magnification. This document is taken from the CNES Web site www.cnes.fr. Protected information. All rights reserved ' CNES (2014).

SPOT 10 × 10 m panchromatic data are of such high geometric fidelity that they can be photointerpreted like a typical aerial photograph in many instances. For this reason, SPOT panchromatic data are often registered to topographic base maps and used as orthophotomaps. Such image maps are useful in GIS databases because they contain more accurate planimetric information (e.g., new roads, subdivisions) than out-of-date 7.5-min topographic maps. The improved spatial resolution available is demonstrated in Figure 2-35, which presents a TM band 3 image and a SPOT panchromatic image of Charleston, SC.

SPOT sensors collect data over a relatively small 60 × 60 km (3,600 km^2) area compared with Landsat MSS and TM image areas of 170 × 185 km (31,450 km^2) (Figure 2-36). It takes 8.74 SPOT images to cover the same area as a single Landsat TM or MSS scene. This may be a limiting factor for regional studies. However, SPOT does allow imagery to be purchased by the km^2 (e.g., for a watershed or school district) or by the linear km (e.g., along a highway).

SPOTs 4 and 5

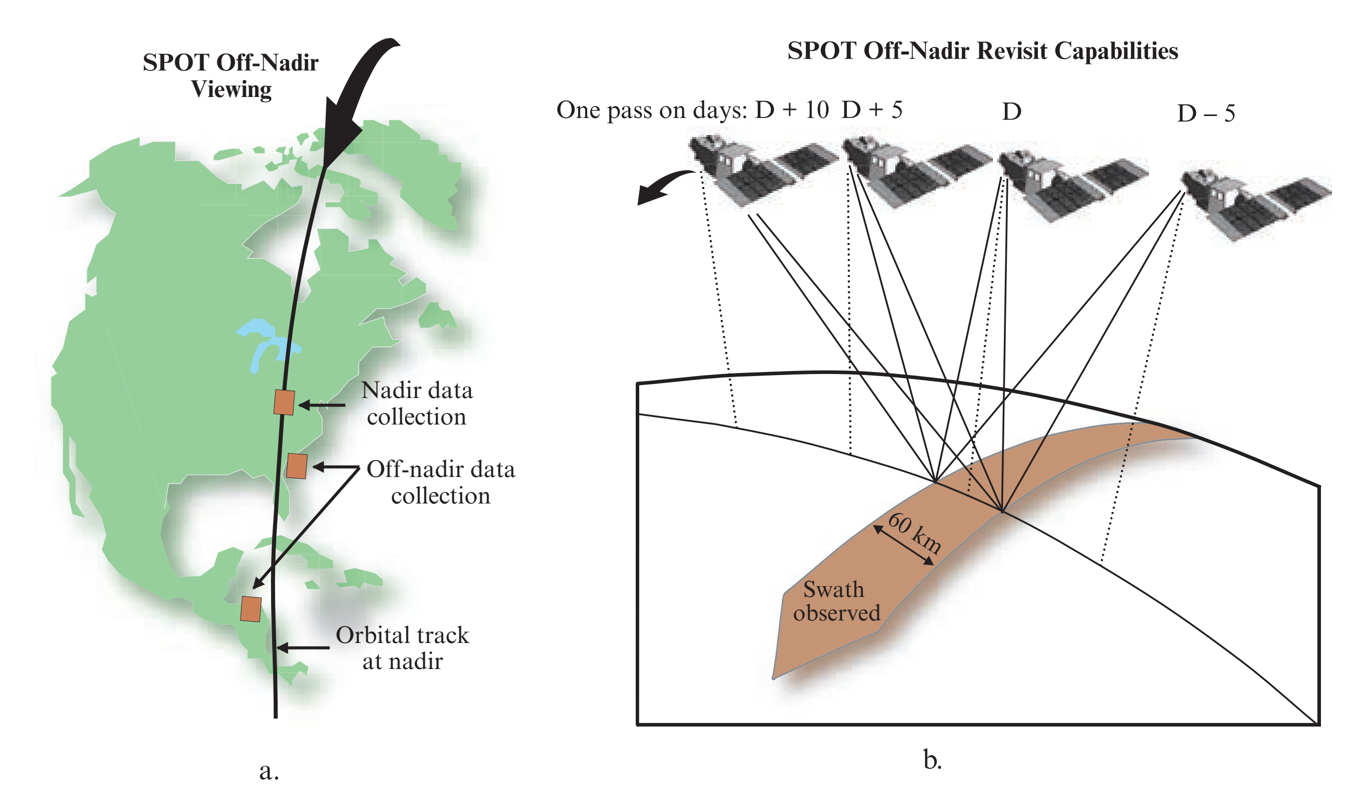
SPOT Image, Inc., launched SPOT 4 on March 24, 1998, and SPOT 5 on May 3, 2002. Their characteristics are summarized in Table 2-10. The viewing angle can be adjusted to ±27° off-nadir. SPOTs 4 and 5 have several notable features of significant value for Earth resource remote sensing, including:

- the addition of a short-wavelength infrared (SWIR) band (1.58 – 1.75 μm) for vegetation and soil moisture applications at 20 × 20 m;
- an independent sensor called *Vegetation* for small-scale vegetation, global change, and oceanographic studies;
- SPOT 4 has onboard registration of the spectral bands, achieved by replacing the original HRV panchromatic sensor (0.51 – 0.73 μm) with band 2 (0.61 – 0.68 μm) operating in both 10- *and* 20-m resolution mode; and
- SPOT 5 panchromatic bands (0.48 – 0.7 μm) can collect 2.5 × 2.5 m imagery.

Because the SPOT 4 and 5 HRV sensors are sensitive to SWIR energy, they are referred to as HRVIR 1 and HRVIR 2.

The SPOT 4 and 5 *Vegetation* sensor is independent of the HRVIR sensors. It is a multispectral electronic scanning radiometer operating at optical wavelengths with a separate objective lens and sensor for each of the four spectral bands (blue = 0.43 – 0.47 μm used primarily for atmospheric correction; red = 0.61 – 0.68 μm; near-infrared = 0.78 – 0.89 μm; and SWIR = 1.58 – 1.75 μm). Each sensor is a 1,728 CCD linear array located in the focal plane of the corresponding objective lens. The spectral resolution of the individual bands are summarized in Table 2-10. The *Vegetation* sensor has a spatial resolution of 1.15 × 1.15 km. The objective lenses offer a field of view of ±50.5°, which translates into a 2,250-km swath width. The *Vegetation* sensor has several important characteristics:

- multidate radiometric calibration accuracy better than 3% and absolute calibration accuracy better than 5% is superior to the AVHRR, making it more useful for repeatable global and regional vegetation surveys;
- pixel size is uniform across the entire swath width, with geometric precision better than 0.3 pixels and



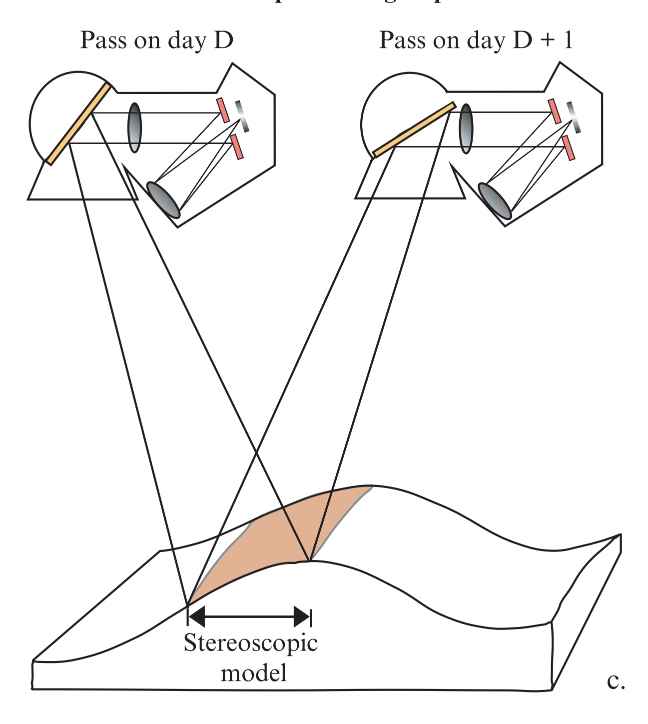


FIGURE 2–34 a) The SPOT HRV instruments are pointable and can be used to view areas that are not directly below the satellite (i.e., off-nadir). This is useful for collecting information in the event of a disaster, when the satellite track is not optimum for collecting stereoscopic imagery. b) During the 26-day period separating two successive SPOT satellite overpasses, a point on Earth could be observed on different passes if it is at the equator and on 11 occasions if at a latitude of 45°. A given region can be revisited on dates separated alternatively by 1, 4, and occasionally 5 days. c) Two observations can be made on successive days such that the two images are acquired at angles on either side of the vertical, resulting in stereoscopic imagery. Such imagery can be used to produce topographic and planimetric maps (adapted from SPOT Image, Inc.).

interband multidate registration better than 0.3 km because of the pushbroom technology;

- 10:30 a.m. equatorial crossing time versus AVHRR's 2:30 p.m. crossing time;
- a short-wavelength infrared band for improved vegetation mapping;
- it is straightforward to relate the HRVIR 10 × 10 m or 20 × 20 m data nested within the *Vegetation* 2,250 × 2,250 km swath width data; and
- individual images can be obtained or data can be summarized over a 24-hr period (called a daily synthesis), or daily synthesis data can be compiled into *n*-day syntheses.

SPOT 4 and 5 are still operational (SPOT Image, 2014; SPOT Payload, 2014).

Comparison of Landsat TM (30 x 30 m) and SPOT HRV (10 x 10 m)



a. Landsat Thematic Mapper Band 3 (30 × 30 m) on February 3, 1994.



b. SPOT HRV Panchromatic Band (10 × 10 m) on January 10, 1996.

FIGURE 2–35 Comparison of the detail in a) 30 × 30 m Landsat TM band 3 data (courtesy of NASA), and b) SPOT 10 × 10 m panchromatic data of Charleston, SC [SPOT image courtesy of CNES (www.cnes.fr). Protected information. All rights reserved '   CNES (2014)].

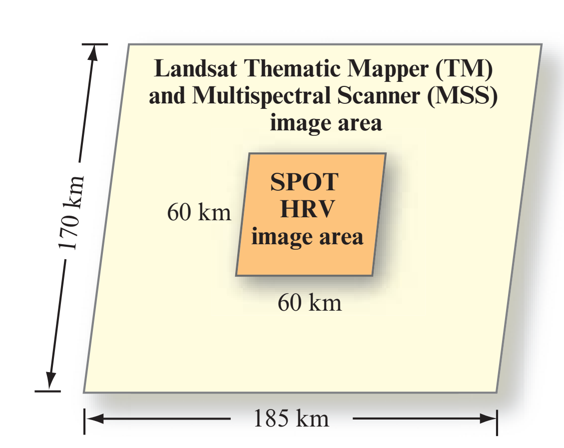


FIGURE 2–36 Geographic coverage of the SPOT HRV and Landsat Multispectral Scanner and Thematic Mapper remote sensing systems.

More than 20 million SPOT images have been collected since 1986.

SPOT 6 and 7

SPOT 6 was launched on September 9, 2012. SPOT 7 was launched on June 30, 2014 (Figure 2-31) (Astrium SPOT 6 and 7, 2014). They are to be placed in phase in the same orbit and are pointable. They will have a 1.5 × 1.5 m panchromatic band and four 6 × 6 m multispectral bands with a 60 km swath width. Both cross-track and along-track stereoscopic data collection is possible. Astrium GEO-Information Services tasks the French SPOT satellite family, Germany's TerraSAR-X, and Taiwan's FORMOSAT-2 directly, enabling rapid and reliable data acquisition of any place on the Earth each day (Astrium, 2014).

Pleiades

The SPOT 6 optical Earth observation satellite combines the advantages of a relatively wide imaging swath and 1.5 × 1.5 m spatial resolution. However, the CNES recognized the need for civilian and military applications that require higher spatial resolution and the collection of data through clouds and at night. Therefore, a joint French-Italian Optical and Radar Federated Earth Observation (ORFEO) space program was initiated. The Pleiades optical system is supervised by the French CNES and the radar imaging system is supervised by the Italian space agency (ISA) (CNES Pleiades, 2014).

Pleiades 1A and 1B

There are two Pleiades satellites. Pleiades 1A was successfully launched on December 17, 2011, into a Sun-synchronous circular orbit at 694 km AGL. Pleiades 1B was launched in 2013 (Figure 2-31). Each has a life span of 5 years. Pleiades satellites record panchromatic imagery (0.48 – 0.83 μm) using linear array technology (Figure 2-5d) at 0.5 × 0.5 m spatial resolution at Nadir (Astrium, 2014). They also record four bands of multispectral data (blue [0.43 – 0.55 μm], green [0.49 – 0.61 μm], red [0.60 – 0.72 μm], and near-infrared [0.75 – 0.95 μm]) at 2 × 2 m spatial resolution with a swath width of 20 km at Nadir (Astrium Pleiades, 2014).

French CNES Pleiades HR-1 Satellite



b. Pleiades HR-1 panchromatic image of San Francisco, CA (0.7 × 0.7 m spatial resolution at nadir).



c. Pleiades HR-1 pansharpened color-composite image of Washington, DC.

FIGURE 2–37 a) Pleiades HR-1 panchromatic image of San Francisco, CA. b) Pansharpened image of Washington, DC (images courtesy of ASTRIUM Geo-Information Services, Inc.; www.astrium-geo.com/).

Merged panchromatic/multispectral images are available as orthophoto products. Pleiades is pointable and can obtain stereoscopic images suitable for photogrammetric analysis to extract detailed planimetric and three-dimensional information. Pleiades has a revisit capability of two days with one satellite and <24 hours with two satellites. Examples of Pleiades imagery of San Francisco, CA, and Washington, DC, are shown in Figure 2-37.

Indian Remote Sensing Systems

The Indian Space Research Organization (ISRO) satellite system is one of the largest and most diverse constellations of remote sensing satellites in operation in the world today (www.isro.org). It began with the launch of the Indian Remote Sensing (IRS) satellites: IRS-1A on March 17, 1988, IRS-1B on August 29, 1991, IRS-1C on December 28,1995, and IRS-1D on

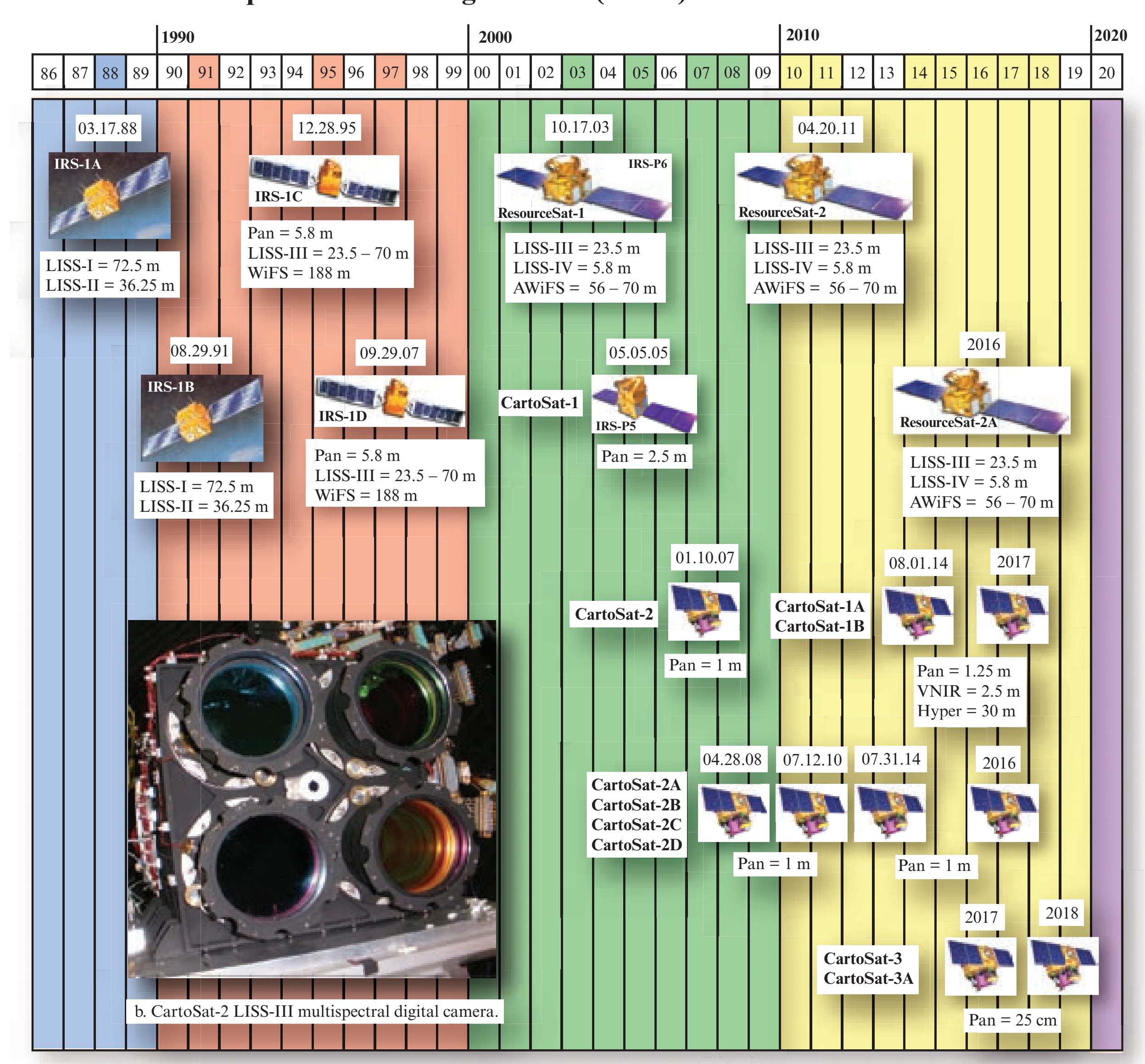


FIGURE 2–38 The chronological launch history of the Indian Space Research Organization (ISRO) optical satellites and their range of spatial resolution (information courtesy of ISRO (www.IRSO.org) and ESA CEOS (2014); Jensen et al., 2012). b) The CartoSat-2 LISS-III multispectral digital camera (National Remote Sensing Centre (NRSC)/Indian Space Research Organization).

September 29, 1997 (Table 2-11). The chronological launch history is shown in Figure 2-38.

IRS-1A, -1B, -1C, and -1D

- The IRS-1A and IRS-1B satellites acquired data with Linear Imaging Self-scanning Sensors (LISS-I and LISS-II) at spatial resolutions of 72.5 × 72.5 m and 36.25 × 36.25 m, respectively. The data were collected in four spectral bands that were nearly identical to the Landsat TM visible and near-infrared bands. The satellite's altitude was 904 km, the orbit was Sun-synchronous, and repeat coverage was every 22 days at the equator (11-day repeat coverage with two satellites), with an orbital inclination of 99.5°. The swath width was 146 to 148 km.
- The IRS-1C and IRS-1D satellites carried three sensors: the LISS-III multispectral sensor, a panchromatic sensor, and a Wide Field Sensor (WiFS). The LISS-III had four bands with the green, red, and near-infrared bands at 23.5 × 23.5 m spatial resolution and the short-wavelength infrared band at 70.5 × 70.5 m spatial resolution. The swath width was

TABLE 2–11 Indian CartoSat and ResourceSat characteristics (ISRO ResourceSat-1, 2014; ISRO ResourceSat-2, 2014).

CartoSat			ResourceSat			
	Spectral Resolution (μm)	Spatial Resolution (m) at Nadir		Band	Spectral Resolution (μm)	Spatial Resolution (m) at Nadir
CartoSat-1 (2 cameras)	0.50 – 0.75	2.5 × 2.5	**ResourceSat-1 ResourceSat-2**			
CartoSat-2	0.50 – 0.75	<1 × 1	**LISS-IV**	2	0.52 – 0.59	5.8 × 5.8
CartoSat-2A	0.50 – 0.75	<1 × 1		3	0.62 – 0.68	5.8 × 5.8
				4	0.77 – 0.86	5.8 × 5.8
			LISS-III	2	0.52 – 0.59	23.5 × 23.5
				3	0.62 – 0.68	23.5 × 23.5
				4	0.77 – 0.86	23.5 × 23.5
				5	1.55 –1.70	23.5 × 23.5
			AWiFS	2	0.52 – 0.59	56 × 56
				3	0.62 – 0.68	56 × 56
				4	0.77 – 0.86	56 × 56
				5	1.55 – 1.70	56 × 56
Sensor	Linear array pushbroom		Linear array pushbroom			
Swath width	CartoSat-1 = 30 km; CartoSat-2 = 9.6 km; CartoSat-2A = 9.6 km; CartoSat-2B= 9.6 km		ResourceSat-1 (LISS-III = 141 km; LISS-IV = 23 km multispectral mode and 70 km mono mode; AWiFS = 740 km) ResourceSat-2 (LISS-III = 141 km; LISS-IV = 70 km; AWiFS = 740 km)			

141 km for bands 2, 3, and 4 and 148 km for band 5. Repeat coverage was every 24 days at the equator.

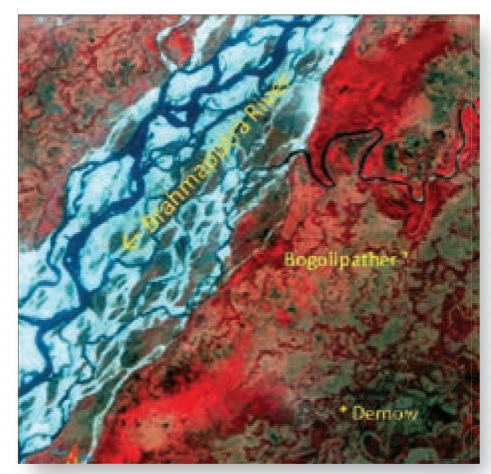
- The panchromatic sensor had a spatial resolution of approximately 5.2 × 5.2 m, stereoscopic imaging capability, and a 70 km swath width with repeat coverage every 24 days at the equator and a revisit time of 5 days with ±26° off-nadir viewing.
- The Wide Field Sensor had 188 × 188 m spatial resolution. The WiFS had two bands comparable to NOAA's AVHRR satellite (0.62 – 0.68 μm and 0.77 – 0.86 μm) with a swath width of 692 km. Repeat coverage was 5 days at the equator.

CartoSat

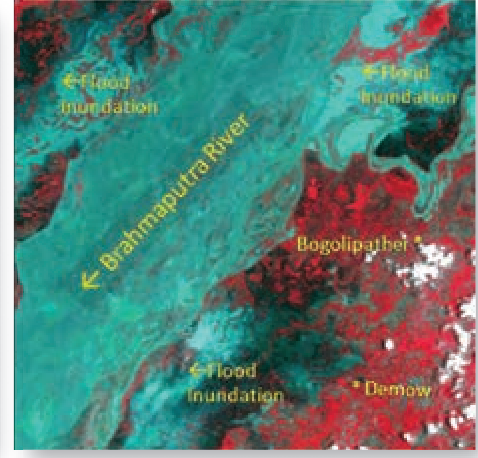
The ISRO has launched numerous high-resolution remote sensing systems that are especially designed for cartographic purposes, including: CartoSat-1, CartoSat-2, CartoSat-2A, and CartoSat-2B. The chronological launch history is shown in Figure 2-38.

CartoSat-1 was launched on May 5, 2005. It carries two panchromatic cameras that have a spatial resolution of 2.5 × 2.5 m with a 30 km swath width (Table 2-11). The cameras are mounted so that near simultaneous imagery of the same geographic area from two different angles is possible yielding stereoscopic imagery. CartoSat-1A launched August 1, 2014, records panchromatic (1.25 × 1.25 m; 500 – 750 nm; 60 km swath width), multispectral VNIR (2.5 × 2.5 m; 60 km swath width), and hyperspectral imagery (30 × 30 m; 60 km swath width; VNIR 750 – 1300 nm; SWIR 30 × 30 m; 1,300 – 3,000 nm). CartoSat-1B may be launched in 2017 with similar instruments (ESA CEOS, 2014).

ResourceSat-2 Imagery of Flooding on the Brahmaputra River in India in 2012




a. Pre-flood ResourceSat-2 (LISS-III) image collected in February, 2012.



b. Post-flood ResourceSat-2 (LISS-III) image collected on June 29, 2012.

FIGURE 2–39 Indian ResourceSat-2 pre- and post-flood imagery of parts of the Bibrugarh, Sibsagar, and Dhemaji Districts in India (images courtesy of Indian Space Research Organization).

CartoSat-2 was launched on January 10, 2007. It carries a single panchromatic camera with sensitivity from 500 – 750 nm that collects data at < 1 × 1 m spatial resolution with a 9.6 km swath width. The camera can be steered up to +45° along-track as well as across-track to obtain stereoscopic imagery. CartoSat-2A was launched on April 28, 2008. Its specifications are almost identical to CartoSat-2. CartoSat-2B was launched on July 12, 2010, into a circular polar Sun-synchronous orbit with 14 orbits per day and a revisit cycle of 4 days. The sensor has a single panchromatic band (500 – 750 nm) at a spatial resolution of < 1 × 1 m (ISRO CartoSat-2B, 2014). It has a swath width of 9.6 km and is steerable up to +26° along-track as well as across-track to obtain stereoscopic imagery and achieve a four-day revisit capability. The sensor is designed for detailed urban and infrastructure planning and development, transportation system planning, and preparation of large-scale cartographic maps (ISRO CartoSat-2B, 2014). According to the European Space Agency (ESA), CartoSat-2C was launched July 31, 2014 with a 1 to 2 m spatial resolution, 10 km swath width, and four bands in the VNIR (400 – 1,300 nm) and CartoSat-2D may be launched in 2016 with characteristics similar to CartoSat-2C (ESA CEOS, 2014). CartoSat-3 to be launched in 2017 will have a panchromatic camera that acquires imagery at 25 × 25 cm spatial resolution in the region from 500 – 750 nm. It will have a swath width of 15 km. CartoSat-3A with similar sensor characteristics is to be launched in 2018 (ESA CEOS, 2014).

ResourceSat

In addition to CartoSat, ISRO has also launched several Earth resource remote sensing satellites, including: ResourceSat-1 and ResourceSat-2. ResourceSat-1 was launched on October 17, 2003, to replace and improve the remote sensing capability of the IRS-1C and IRS-1D Earth resource satellites (Figure 2-38). It carried three cameras including the high-resolution Linear Imaging Self Scanner (LISS-IV) with three bands at 5.8 × 5.8 m spatial resolution, steerable +26° across-track viewing to achieve five-day revisit capability; a LISS-III sensor operating in three spectral bands in the VNIR and one in the Short Wave Infrared (SWIR) band with 23.5 × 23.5 spatial resolution; and a coarse-resolution Advance Wide Field Sensor (AWiFS) operating in three spectral bands in VNIR and one band in SWIR at 56 × 56 m. It has a revisit capability of 5 days. ResourceSat-2 was launched on April 20, 2011. It carries the three cameras that are similar to those of ResourceSat-1, including the LISS-IV, LISS-III (Figure 2-38), and the AWiFS. Important changes in ResourceSat-2 compared to ResourceSat-1 include: enhancement of LISS-IV multispectral swath from 23 km to 70 km and improved radiometric accuracy from 7-bits to 10-bits for LISS-III and LISS-IV and 10-bits to 12-bits for AWIFS. A follow-on satellite ResourceSat-2A is scheduled for launch in 2016 (ESA CEOS, 2014). Examples of pre- and post-flood ResourceSat-2 imagery of an area in India are shown in Figure 2-39.

Korean Aerospace Research Institute (KARI) KOMPSATs

The Korea Aerospace Research Institute (KARI) supervises Korea's satellite programs (GlobalSecurity.org). The KOMPSAT (Korea Multi-Purpose Satellite) program was initiated in 1995 (Figure 2-40). Its objective is the development of a national space

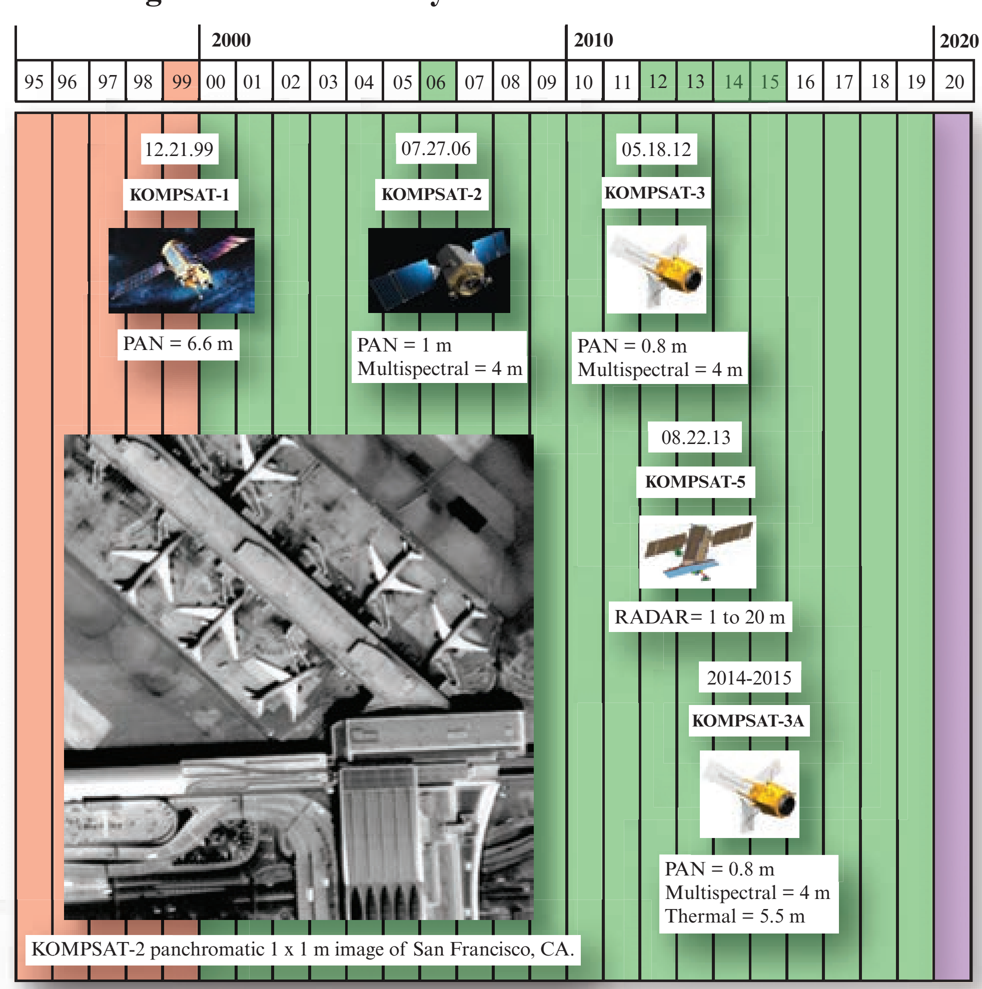


FIGURE 2–40 The chronological launch history of the Korean Aerospace Research Institute (KARI) satellites and their spatial resolutions. A KOMPSAT-2 panchromatic 1 × 1 m image of San Francisco International Airport obtained in 2006 is shown (image ' Korean Aerospace Research Institute (KARI), 2006; information courtesy of KARI and SPOT Image, Inc.; Jensen et al., 2012).

segment in Earth observation along with an efficient infrastructure and ground segment to provide services to remote sensing users.

KOMPSAT-1 (Arirang-1) was launched on December 21, 1999, and retired on January 1, 2008. The payload included a CCD pushbroom Electro-Optical Camera (EOC) with a 6.6 × 6.6 m spatial resolution in the region from 510 – 730 nm. It had a 28 revisit cycle and a swath width of 17 km (Figure 2-40).

KOMPSAT-2 (Arirang-2) was launched on July 27, 2006, and orbits the Earth 14 times a day. It has a 1 × 1 m panchromatic sensor (500 – 900 nm), a 4 × 4 m four-band VNIR multispectral camera, and a 15 km swath width (Figure 2-40). KARI developed the KOMPSAT-2 program in collaboration with EADS Astrium, Inc., to acquire VHR imagery for South Korea's mapping requirements, urban planning, and hazard management. A KOMPSAT-2 image of San Francisco International Airport is shown in Figure 2-40.

KOMPSAT-3 was launched on May 18, 2012. It has an Advanced Electronic Image Scanning System (AEISS) with a 0.8 × 0.8 m panchromatic band and a 4 × 4 m VNIR multispectral sensor system (Figure 2-40). The swath width is 15 km.

KOMPSAT-3A may be launched in 2014-2015. It will have panchromatic and multispectral remote sensing capability similar to KOMPSAT-3 except for a thermal infrared band in the 3 – 6 μm region at a spatial resolution of 5.5 × 5.5 m. It will have a 15 km swath width (ESA CEOS, 2014). There is no KOMPSAT-4 program. The Sino-Korean word for the number four, "sa" is a homonym of the Chinese character for death.

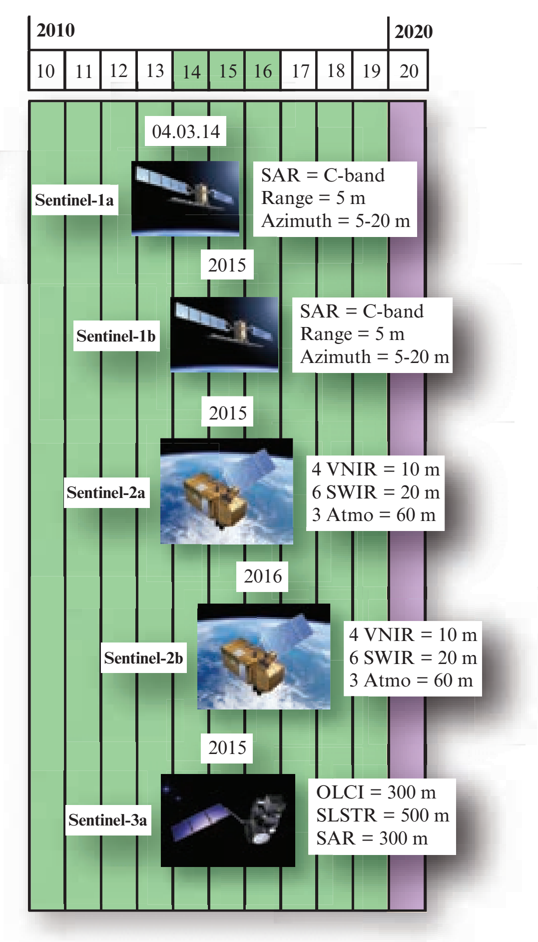


FIGURE 2–41 Proposed launch history of the Global Monitoring for Environment and Security (GMES) program European Space Agency Sentinel 1-3 satellites (information courtesy of ESA CEOS, 2014; updated from Jensen et al., 2012).

KOMPSAT-5 provides high-resolution Synthetic Aperture Radar (SAR) images at 1 × 1 m spatial resolution, standard mode SAR images of 3 × 3 resolution, and wide swath mode SAR images of 20 × 20 m resolution with an incidence angle of 45° using the COSI (COrea SAR Instrument) payload. KOMPSAT-5 was launched August 22, 2013. It has an X-band (12.5 – 8 GHz) Synthetic Aperture Radar (SAR) with a 100 km swath width and a 28-day revisit cycle (ESA CEOS, 2014; eoPortal KOMPSAT-5, 2014) (Figure 2-40). KOMPSAT-6 will be a SAR and is scheduled for launch in 2019.

Astrium, Inc. Sentinel-2

The Global Monitoring for Environment and Security (GMES) program is headed by the European Commission (EC) in partnership with the European Space Agency (ESA) and the European Environment Agency (EEA). The ambitious Earth observation program is to provide accurate, timely and easily accessible information to improve the management of the environment, understand and mitigate the effects of climate change and ensure civil security. Sentinel-1a is a C-band radar satellite launched on April 3, 2014 (Figure 2-41). Sentinel-2 records high-resolution optical multispectral imagery for land applications. Sentinel-3 will provide data for ocean and land applications. Sentinel-4 and Sentinel-5 will provide data for atmospheric composition monitoring from geostationary and polar orbits, respectively (ESA GMES Sentinel-2, 2014; ESA Sentinel-2, 2014). Sentinel-2 services fall into six main categories: services for land management, services for the marine environment, services relating to the atmosphere, services to aid emergency response, services associated with security, and services relating to climate change. The mission will adhere to the Sentinel Data Policy, which establishes full and open access to data acquired by all of the Sentinels.

A pair of Sentinel-2(ab) satellites will deliver high-resolution optical images globally, providing enhanced continuity of SPOT and Landsat-type data. The first satellite is scheduled to launch in 2015 and the next within 18 months. Astrium GmbH is the prime contractor and has responsibility for the design, development, and the integration of the Sentinel-2 satellite (Figure 2-41). Sentinel-2 will carry a Multi-Spectral Instrument (MSI) with a total of 13 bands: 4 VNIR bands (400 – 750 nm) at 10 m, 6 SWIR bands (1,300 – 3,000 nm) at 20 m and 3 bands at 60 m spatial dedicated to atmospheric corrections and cloud screening. It has a swath width of 290 km (ESA CEOS, 2014). The mission orbits at a mean altitude of approximately 800 km and, with the pair of satellites in operation, will have a revisit time of 5 days at the equator (under cloud-free conditions) and 2–3 days at mid-latitudes. The increased swath width along with the short revisit time allow rapid changes to be monitored, such as vegetation during the growing season. Data from Sentinel-2 will benefit services associated with land management by European and national institutes, the agricultural industry and forestry, as well as disaster control and humanitarian relief operations.

Imagery for the generation of high-level operational products, such as land-cover maps, land-change detection maps and geophysical variables that use, for example, leaf-area-index, leaf chlorophyll content and leaf water content will be provided. Images of floods, volcanic eruptions and landslides will also be acquired by Sentinel-2. In essence, Sentinel-2 combines a large swath, frequent revisit, and systematic acquisition of

TABLE 2–12 NASA Advanced Spaceborne Thermal Emission and Reflection Radiometer characteristics.

Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)					
Band	**VNIR Spectral Resolution (μm)**	**Band**	**SWIR Spectral Resolution (μm)**	**Band**	**TIR Spectral Resolution (μm)**
1 (nadir)	0.52 – 0.60	4	1.600 – 1.700	10	8.125 – 8.475
2 (nadir)	0.63 – 0.69	5	2.145 – 2.185	11	8.475 – 8.825
3 (nadir)	0.76 – 0.86	6	2.185 – 2.225	12	8.925 – 9.275
3 (backward)	0.76 – 0.86	7	2.235 – 2.285	13	10.25 – 10.95
		8	2.295 – 2.365	14	10.95 – 11.65
		9	2.360 – 2.430		
Technology (detector)	Pushbroom Si		Pushbroom PtSi:Si		Whiskbroom Hg:Cd:Te
Spatial resolution (m)	15 × 15		30 × 30		90 × 90
Swath width	60 km		60 km		60 km
Quantization	8 bits		8 bits		12 bits

all land surfaces at high spatial resolution and with a large number of spectral bands, all of which makes a unique mission to serve GMES.

Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)

The ASTER is a joint effort between NASA and Japan's Ministry of International Trade and Industry. ASTER collects information on surface temperature, emissivity, reflectance, and elevation (NASA ASTER, 2014a). It is the only high-to-moderate-spatial-resolution instrument on the *Terra* satellite. It is used in conjunction with MODIS, MISR, and CERES sensors that monitor the Earth at moderate to coarse spatial resolutions. ASTER serves as a zoom lens for the other *Terra* instruments and is particularly important for change detection and calibration/validation studies.

ASTER collects data in 14 channels from the visible through the thermal-infrared regions of the electromagnetic spectrum. The sensor system consists of three separate instrument subsystems. The individual bandwidths and subsystem characteristics are summarized in Table 2-12.

The VNIR detector subsystem operates in three spectral bands in the visible and near-infrared wavelength region with a spatial resolution of 15 × 15 m. It consists of two telescopes—one nadir-looking with a three-spectral-band CCD detector and another backward-looking with a single-band CCD detector. The backward-looking telescope provides a second view of the study area in band 3 for stereoscopic observations. Across-track pointing to 24° off-nadir is accomplished by rotating the entire telescope assembly. ASTER optical VNIR imagery of Waikiki, HI, is shown in Figure 2-42.

The SWIR subsystem collects data in six spectral bands in the 1.6 – 2.43 μm region through a Nadir-pointing telescope that provides 30 × 30 m spatial resolution. Cross-track pointing (± 8.55°) is accomplished by a pointing mirror.

The TIR subsystem operates in five bands in the thermal-infrared region using a single nadir-looking telescope with a spatial resolution of 90 × 90 m. Unlike the other subsystems, it has a whiskbroom scanning system. Each band uses 10 detectors with optical band-pass filters over each detector element. The scanning mirror functions both for scanning and cross-track

Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Optical Imagery

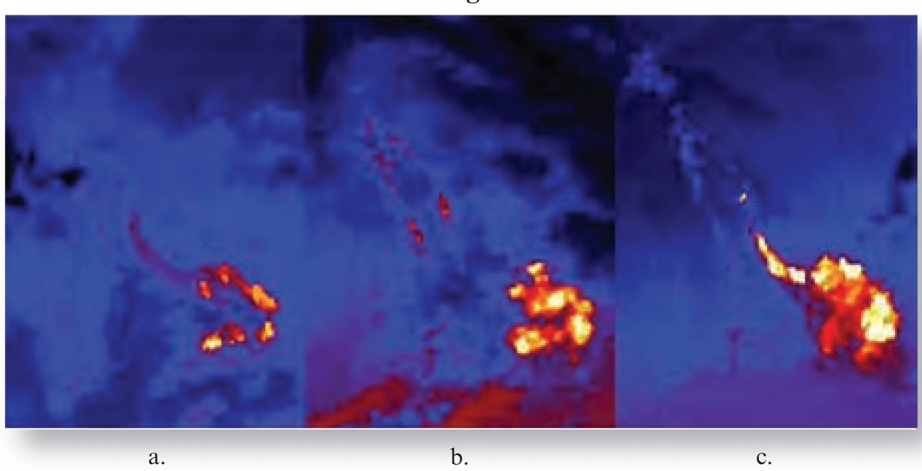


a. Visible near-infrared (VNIR) image of Waikiki, HI, (15 × 15 m spatial resolution; RGB = bands 2,3,1).

b. 3× magnification.

FIGURE 2–42 *Terra* ASTER imagery of Waikiki and Diamond Head on the island of Oahu, Hawaii (courtesy of NASA/GSFC/MITI/ERSADC/JAROS and U.S./Japan ASTER Science Team and the California Institute of Technology).

ASTER Thermal-infrared Images of a Lava Flow on Hawaii



a. b. c.

FIGURE 2–43 A sequence of *Terra* ASTER nighttime 90 × 90 m thermal infrared band 14 (10.95 – 11.65 μm) images of Pu'u O'o lava flows entering the sea at Kamokuna on the southeast side of the Island of Hawaii. The images were obtained on a) May 22, 2000, b) June 30, 2000, and c) August 1, 2000 (courtesy of NASA/GSFC/MITI/ERSADC/JAROS and U.S./Japan ASTER Science Team and the California Institute of Technology).

pointing (± 8.55°). During scanning, the mirror rotates 90° from nadir to view an internal blackbody. Multi-date nighttime 90 × 90 m thermal infrared band 14 (10.95 – 11.65 μm) images of Pu'u O'o lava flows entering the sea at Kamokuna on the southeast side of the Island of Hawaii are shown in Figure 2-43.

Multiangle Imaging Spectroradiometer (MISR)

The Multiangle Imaging Spectroradiometer was built by NASA's Jet Propulsion Laboratory and is one of the five *Terra* satellite instruments. MISR measures the Earth's brightness in four spectral bands, at each of nine look angles spread out in the forward and aft directions along the flight line. Spatial samples are acquired every 275 m. Over a period of 7 minutes, a 360-km wide swath of Earth comes into view at all nine angles (NASA MISR, 2014).

An illustration of the nine look angles is shown in Figure 2-44. The digital pushbroom sensors image the Earth at 26.1°, 45.6°, 60°, and 70.5° forward and aft of the local vertical (Nadir 0°). Note that the fore and aft camera angles are arranged symmetrically about nadir. In general, large viewing angles provide enhanced sensitivity to atmospheric aerosol effects and to cloud reflectance effects, whereas more modest angles are required for land-surface viewing.

Each MISR camera sees instantaneously a single row of pixels in a pushbroom format. It records data in

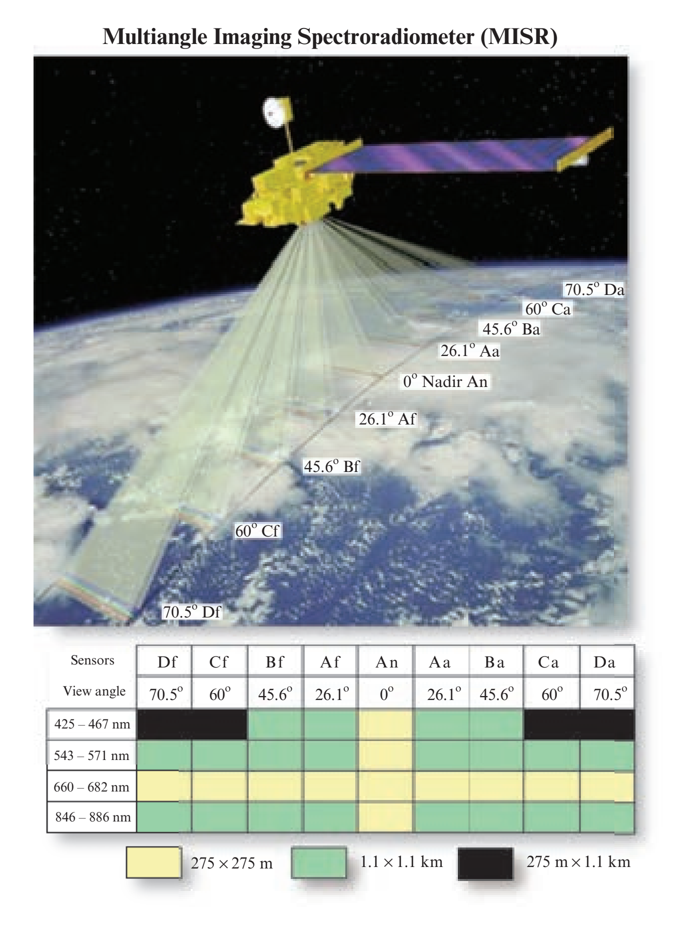


FIGURE 2–44 Artist's rendition of the Multiangle Imaging Spectroradiometer (MISR) on EOS *Terra*. MISR uses linear array technology to acquire imagery of the terrain in four bands at nine angles: at Nadir (0°) and at 26.1°, 45.6°, 60°, and 70.5° forward and aft of nadir (adapted from NASA Jet Propulsion Laboratory).

four bands: blue, green, red, and near-infrared. The individual band wavelengths are summarized in Figure 2-44. Each camera has four independent linear CCD arrays (one per filter), with 1,504 active pixels per linear array.

The nadir-viewing camera (labeled *An* in Figure 2-44) provides imagery that is less distorted by surface topographic effects than that of any other MISR camera. It also is the least affected by atmospheric scattering. It provides 1) useful reference for navigating within all the MISR imagery, and 2) a base image to compare with images acquired at different angles of view. Such comparisons provide important "bidirectional reflectance distribution function, BRDF" information introduced in Chapter 1. The nadir-viewing camera also offers an opportunity to compare observations with other nadir-viewing sensors such as Landsat TM and ETM$^+$. The nadir-viewing camera also facilitates calibration.

The fore and aft 26.1° view angle cameras (*Af* and *Aa*) provide useful stereoscopic information that can be of benefit for measuring topographic elevation and cloud heights. The fore and aft 45.6° view angle cameras (*Bf* and *Ba*) are positioned to be especially sensitive to atmospheric aerosol properties. The fore and aft 60° view angle cameras (*Cf* and *Ca*) provide observations looking through the atmosphere with twice the amount of air than the vertical view. This provides unique information about the hemispherical albedo of land surfaces. The fore and aft 70.5° view angle cameras (*Df* and *Da*) provide the maximum sensitivity to off-nadir effects. The scientific community is interested in obtaining quantitative information about clouds and the Earth's surface from as many angles as possible.

GeoEye, Inc. (formerly Space Imaging, Inc.), IKONOS-2, GeoEye-1, GeoEye-2

In 1994, the U.S. government made a decision to allow civil commercial companies to market high spatial resolution remote sensor data (approximately 1 × 1 to 4 × 4 m). This resulted in the creation of a number of commercial consortiums that had the capital necessary to create, launch, and market high spatial resolution digital remote sensor data. The most notable companies are Space Imaging, Inc., Orbimage, Inc., and EarthWatch, Inc. These companies targeted the geographic information system (GIS) and cartographic mapping markets traditionally serviced by the aerial photogrammetric industries. The commercial remote sensing firms hoped to have an impact in markets as diverse as agriculture, natural resource management, local and regional government, transportation, emergency response, mapping, and eventually an array of average consumer applications such as *Google Earth*. Warfare during the early twenty-first century also created a significant military demand for commercial high spatial resolution imagery.

All commercial vendors offer an Internet online ordering service and a suite of products that can be tailored to user requirements, including the creation of digital elevation models from the remote sensor data. The commercial remote sensing companies typically price the imagery according to the type of product ordered and the amount of geographic coverage desired (km^2). The sensors used by these companies are based primarily on linear array CCD technology.

IKONOS-1 and -2

Space Imaging, Inc. launched IKONOS-1 on April 27, 1999. Unfortunately, the IKONOS-1 satellite never achieved orbit. Space Imaging, Inc. successfully launched the IKONOS-2 satellite on September 24, 1999 (Figure 2-45). The IKONOS-2 sensor system has

Chronological Launch History of the IKONOS and GeoEye Satellites

09.24.99 IKONOS-2 — PAN = 0.82 m, Multispectral = 3.2 m

09.06.08 GeoEye-1 — PAN = 0.41 m, Multispectral = 1.65 m

Future GeoEye-2 — PAN = 0.36 m, Multispectral = 1.65 m

IKONOS-2 image of Minot, ND, obtained on May 30, 2009, at 1 x 1 m spatial resolution.

GeoEye-1 image of Minot, ND, obtained on June 25, 2011, at 0.5 x 0.5 m spatial resolution.

FIGURE 2–45 The chronological launch history of the IKONOS and GeoEye satellites and their range of spatial resolution. IKONOS-2 and GeoEye-1 images of Minot, ND, are shown (images courtesy of DigitalGlobe, Inc.; Jensen et al., 2012).

a 1 × 1 m panchromatic band (Figure 2-45) and four multispectral visible and near-infrared bands at 4 × 4 m spatial resolution. Sensor characteristics are summarized in Table 2-13. IKONOS-2 is in a Sun-synchronous 681-km orbit, with a descending equatorial crossing time of between 10 and 11 a.m. It has both cross-track and along-track viewing instruments, which enable flexible data acquisition and frequent revisit capability: <3 days at 1 × 1 m spatial resolution (for look angles < 26°) and 1.5 days at 4 × 4 m spatial resolution. The nominal swath width is 11 km. Data are quantized to 11-bits.

GeoEye-1 and -2

GeoEye, Inc. was established in 2006 as a result of the merger of Orbital Imaging, Inc. and Space Imaging, Inc. It launched GeoEye-1 on September 6, 2008. It has a 41 × 41 cm panchromatic band and four multispectral bands at 1.65 × 1.65 m spatial resolution (Table 2-13). IKONOS-2 and GeoEye-1 images are shown in Figure 2-45. GeoEye, Inc. was merged into DigitalGlobe, Inc. on January 29, 2013. GeoEye-2 (to be known as WorldView-4) is to be launched in 2016. It will have 0.34 × 0.34 m spatial resolution, four multispectral bands, revisit frequency of < 3 days, 14 km swath width, and positional accuracy of < 5 m CE90 (GeoEye, 2014). Until recently, the Federal government only permitted distribution of imagery at a resolution of better 50 × 50 cm outside of the U.S. government and its cooperators (USGS, 2014).

The Federal government Enhanced View program was a $7.3 billion, 10-year contract for commercial satellite imagery from two publicly traded U.S. companies: GeoEye, Inc. and DigitalGlobe, Inc. These companies invested >$1 billion in new satellite remote sensing infrastructure based on this government program, and the majority of their revenues come from government contracts.

EarthWatch/DigitalGlobe, Inc., QuickBird, WorldView-1, WorldView-2, WorldView-3

EarthWatch, Inc. was one of the pioneers in commercial high spatial resolution satellite remote sensing.

TABLE 2–13 Sensor characteristics of satellites operated by GeoEye, Inc., which has merged with DigitalGlobe, Inc.

IKONOS-2 (1999)			GeoEye-1 (2008)			GeoEye-2 (WorldView-4)		
Band	Spectral Resolution (nm)	Spatial Resolution (m) at Nadir	Band	Spectral Resolution (nm)	Spatial Resolution (m) at Nadir	Band	Spectral Resolution (nm)	Spatial Resolution (m) at Nadir
1	445 – 516	3.2 × 3.2	1	450 – 510	1.65 × 1.65	1	450 – 510	1.36 × 1.36
2	506 – 595	3.2 × 3.2	2	510 – 580	1.65 × 1.65	2	510 – 580	1.36 × 1.36
3	632 – 698	3.2 × 3.2	3	655 – 690	1.65 × 1.65	3	655 – 690	1.36 × 1.36
4	757 – 853	3.2 × 3.2	4	780 – 920	1.65 × 1.65	4	780 – 920	1.36 × 1.36
Pan	526 – 929	0.82 × 0.82	Pan	450 – 800	0.41 × 0.41	Pan	450 – 510	0.34 × 0.34
Sensor	Linear array pushbroom		Linear array pushbroom			Linear array pushbroom		

QuickBird

EarthWatch, Inc., launched EarlyBird in 1996 with a 3 × 3 m panchromatic band and three visible to near-infrared (VNIR) bands at 15 × 15 m spatial resolution. Unfortunately, EarthWatch, Inc. lost contact with the satellite shortly after launch. DigitalGlobe, Inc. launched QuickBird on October 18, 2001, with a 65 × 65 cm panchromatic band and four 2.62 × 2.62 m multispectral bands (Figure 2-46; Table 2-14). The QuickBird orbit was recently raised to 482 km AGL which results in a slight decrease in spatial resolution (DigitalGlobe QuickBird, 2014; USGS, 2013).

WorldView-1, -2, and -3

DigitalGlobe, Inc. launched WorldView-1 on September 18, 2007, with a 0.5 × 0.5 m panchromatic band (DigitalGlobe WorldView-1, 2014). WorldView-2 was launched on October 8, 2009, with a 46 × 46 cm panchromatic band and eight multispectral bands at 1.85 × 1.85 m spatial resolution (Table 2-14) (DigitalGlobe WorldView-2, 2014; USGS, 2013). DigitalGlobe launched WorldView-3 on August 13, 2014. It has one panchromatic band (450 – 800 nm) at 31 × 31 cm spatial resolution, eight multispectral bands at 1.24 × 1.24 m spatial resolution, and 8 SWIR bands with 3.7 × 3.7 m resolution (DigitalGlobe WorldView-3, 2014).

ImageSat International, Inc., EROS A and EROS B

ImageSat International, Inc. is a commercial provider of high-resolution, satellite Earth-imagery collected by its Earth Remote Observation Satellites (EROS) (ImageSat EROS, 2014).

EROS A and EROS B

ImageSat, Inc. successfully launched its first satellite, EROS A, on December 5, 2000. In so doing, ImageSat became the second company in the world to successfully deploy a non-government-owned high-resolution imaging satellite. ImageSat, Inc. successfully launched its second satellite, EROS B, on April 25, 2006 (Figure 2-47).

The EROS satellites acquire high-resolution panchromatic (500 – 900 nm) image data using pushbroom linear array technology at 10-bit radiometric resolution. The satellites are deployed in a Sun-synchronous near-polar orbit at an altitude of approximately 500 km. EROS A and B satellites orbit the earth approximately 15 times each day delivering imagery data in real-time to ground receiving stations worldwide. EROS A and B sensors can be pointed and stabilized to image customer specified sites at nadir (perpendicular to the surface) or at oblique angles up to 45°. Oblique viewing enables the satellite to view any site on the Earth as often as two to three times per week.

EROS A is equipped with a camera whose focal plane of CCD (Charge-Coupled Device) detectors produces a standard image spatial resolution of 1.9 × 1.9 m with a swath of 14 km at nadir. The EROS B satellite has a larger panchromatic band camera of CCD/TDI type (Charge-Coupled Device/Time Delay Integration) with a spatial resolution of 0.70 × 0.70 m, a larger onboard recorder, improved pointing accuracy and a faster data communication link. The expected lifespan of both satellites is 14 years. Recently, ImageSat, Inc. and RapidEye, Inc. reached an agreement whereby EROS panchromatic data can be fused with RapidEye multispectral data (ImageSat EROS, 2014).

FIGURE 2–46 The chronological launch history of the QuickBird, WorldView-1, -2, and -3. A WorldView-2 image of San Diego, CA, is shown (images courtesy of DigitalGlobe, Inc.; updated from Jensen et al., 2012).

RapidEye, Inc.

The RapidEye constellation of five identical Earth observation satellites was launched on August 29, 2008, (using a refurbished ICBM missile) and became operational in February, 2009 (Figure 2-48). The constellation acquires over 4 million km^2 of Earth imagery each day and has amassed more than 2.5 billion km^2 of multispectral imagery.

RapidEye

The RapidEye satellite was launched on August 29, 2008. The RapidEye multispectral pushbroom sensor has five bands at 5 × 5 m spatial resolution, a swath width of 77 km, with a daily revisit time. RapidEye's satellites are the first commercial satellites to offer the Red-edge band (690 – 730 nm), which measures variances in vegetation allowing for species separation and monitoring vegetation health (Figure 2-48). The Red-edge is of value for discriminating between urban and vegetation land cover (Dykstra, 2012; RapidEye, 2014). RapidEye level 1b is radiometrically corrected only. The RapidEye Ortho Product has radiometric, sensor, and geometric corrections applied to the data and is rectified using DTED Level 1 SRTM DEM or better. With appropriate ground control it can meet an accuracy of 6 m (1 sigma or 12.7 m CE90) meeting 1:25,000 NMAS standards (RapidEye, 2014). RapidEye, Inc., imagery is distributed in the U.S. through Blackbridge, Inc. (http://blackbridge.com), and Photo Science (Raber, 2012).

DMC International Imaging Ltd. SLIM-6 and NigeriaSat-2

The DMC Consortium satellites are owned and operated by different nations in the consortium. They are placed in phased Sun-synchronous orbits to achieve daily revisit capability. The consortium, established in 2003, has regular additions of new member satellites.

SLIM-6

The DMC SLIM-6 linear array pushbroom sensor system has a 22 × 22 m spatial resolution with three multispectral bands (green, red, and near-infrared) that are Landsat ETM$^+$ equivalent. It is a medium-resolution remote sensing system that is cross-calibrated against a

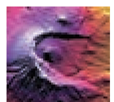
TABLE 2–14 Sensor characteristics of DigitalGlobe, Inc., QuickBird satellite; WorldView-1 and WorldView-2 satellites.

DigitalGlobe, Inc. QuickBird (2001)			WorldView-1 (2007)			WorldView-2 (2009)		
Band	Spectral Resolution (nm)	Spatial Resolution (m) at Nadir	Band	Spectral Resolution (μm)	Spatial Resolution (m) at Nadir	Band	Spectral Resolution (μm)	Spatial Resolution (m) at Na-dir
1	430 – 545	2.62 × 2.62	1	397 – 905	0.5 × 0.5	1	396 – 458	1.85 × 1.85
2	466 – 620	2.62 × 2.62				2	442 – 515	1.85 × 1.85
3	590 – 710	2.62 × 2.622				3	506 – 586	1.85 × 1.85
4	715 – 918	2.62 × 2.62				4	584 – 632	1.85 × 1.85
Pan	405 – 1053	0.65 × 0.65				5	624 – 694	1.85 × 1.85
						6	699 – 749	1.85 × 1.85
						7	765 – 901	1.85 × 1.85
						8	856 – 1043	1.85 × 1.85
						Pan	447 – 808	0.46 × 0.46
Sensor	Linear array pushbroom		Linear array pushbroom			Linear array pushbroom		

reference standard satellite (Landsat 7) over the Libya 4 site and Dome C Antarctica (DMC, 2014a).

DMC-NigeriaSat-2

DMC NigeriaSat-2 has a 2.5 × 2.5 m panchromatic band and four multispectral bands at 5 × 5 m spatial resolution. The swath width is 20.5 km (DMC, 2014b). It is pointable with a 2-day revisit to anywhere on the globe.



Imaging Spectrometry Using Linear and Area Arrays

In the past, most remotely sensed data were acquired in 4 to 12 spectral bands. Imaging spectrometry makes possible the acquisition of data in hundreds of spectral bands simultaneously. Because of the very precise nature of the data acquired by imaging spectrometry, more Earth resource problems can be addressed in greater detail.

The value of an imaging spectrometer lies in its ability to provide a high-resolution reflectance spectrum for each picture element in the image. The reflectance spectrum in the region from 0.4 – 2.5 μm can be used to identify a large range of surface cover materials that cannot be identified with broadband, low spectral resolution imaging systems such as the Landsat MSS, TM, and SPOT. Many surface materials have diagnostic absorption features that are only 10 to 30 nm wide. Therefore, spectral imaging systems that acquire data in contiguous 10-nm bands may produce data with sufficient resolution for the direct identification of those materials with diagnostic spectral absorption features. For example, Figure 2-49 depicts high spectral resolution crop spectra over the interval 400 to 1,000 nm obtained using an Imaging Spectrometer for an agricultural area near Bakersfield, CA. The absorption spectra for the Pima and Royale cotton differ from one another from about 725 nm, where the "red edge" is located, to about 900 nm, leading to the possibility that species within the same crop type might be distinguishable. The Landsat and SPOT sensors, which have relatively large bandwidths, may not be able to resolve these spectral differences.

Two approaches to imaging spectrometry are shown in Figures 2-9ef. The whiskbroom scanner linear array approach (Figure 2-9e) is analogous to the scanner approach used for ETM^+, except that radiant flux from within the IFOV is passed on to a spectrometer, where it is dispersed and focused onto a linear array of detectors. Thus, each pixel is simultaneously sensed in as

ImageSat International EROS B Panchromatic Imagery



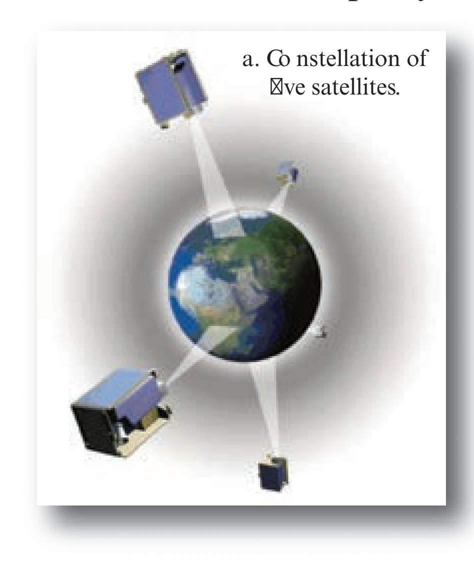
a. ImageSat EROS B satellite.



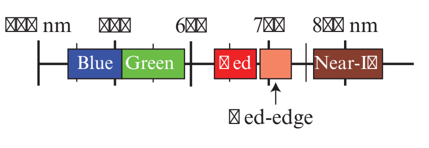
b. ImageSat EROS B 0.7 × 0.7 m panchromatic image of a bridge collapse in Minneapolis, MN.

FIGURE 2–47 a) The ImageSat International EROS B satellite collects panchromatic (500 – 900 nm) imagery at 0.7 × 0.7 m spatial resolution. b) EROS B image of a bridge collapse in Minneapolis, MN (images courtesy of ImageSat International, NV).

RapidEye™ Multispectral Satellite Imagery



a. Constellation of five satellites.



b. Spectral resolution.



c. Image of San Diego, CA, obtained on June 18, 2011.

FIGURE 2–48 a) The constellation of five RapidEye satellites. b) Spectral resolution of the RapidEye sensors including the special red-edge band. c) RapidEye image of San Diego, CA, obtained on June 18, 2011 (images courtesy of BlackBridge).

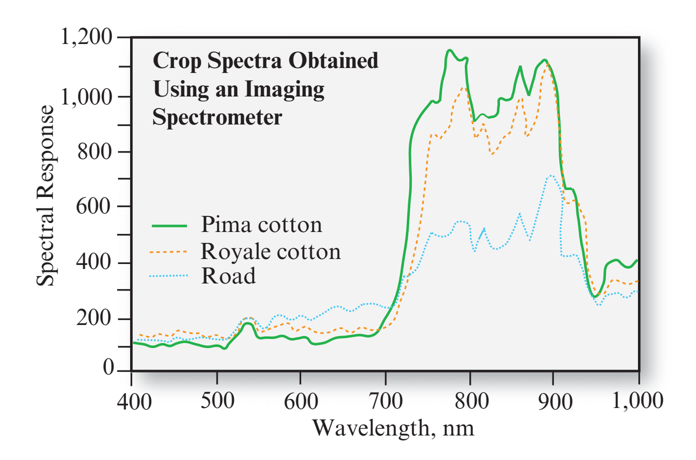


FIGURE 2–49 Imaging spectrometer crop spectra for Pima cotton, Royale cotton, and road surface extracted from 2 × 2 m data obtained near Bakersfield, CA.

many spectral bands as there are detector elements in the linear array. Because of high spacecraft velocities, orbital imaging spectrometry might require the use of two-dimensional area arrays (Figure 2-9f). This eliminates the need for the optical scanning mechanism. In this situation, there is a dedicated column of spectral detector elements for each linear array cross-track pixel in the scene.

Thus, traditional broadband remote sensing systems such as Landsat MSS and SPOT HRV *undersample* the information available from a reflectance spectrum by making only a few measurements in spectral bands up to several hundred nanometers wide. Conversely, imaging spectrometers sample at close intervals (bands on the order of tens of nanometers wide) and have a sufficient number of spectral bands to allow construction of spectra that closely resemble those measured by laboratory instruments. Analysis of imaging spectrometer data allows extraction of a detailed spectrum for each picture element in the image (Figure 2-49). Such spectra often allow direct identification of specific materials within the IFOV of the sensor based upon their reflectance characteristics, including minerals, atmospheric gases, vegetation, snow and ice, and dissolved matter in water bodies.

Analysis of hyperspectral data often requires the use of sophisticated digital image processing software (discussed in Chapter 11). This is because it is usually necessary to calibrate (convert) the raw hyperspectral radiance data to scaled reflectance before it can be properly interpreted. This means removing the effects of atmospheric attenuation, topographic effects (slope, aspect), and any sensor anomalies. Similarly, to get the most out of the hyperspectral data it is usually necessary to use algorithms that 1) allow one to analyze a typical spectra to determine its constituent materials, and 2) compare the spectra with a library of spectra obtained using handheld spectroradiometers such as those provided by the U.S. Geological Survey.

Government agencies and commercial firms have designed hundreds of imaging spectrometers capable of acquiring hyperspectral data. It is beyond the scope of this book to list them all. A few satellite hyperspectral instruments are briefly discussed, including: NASA's EO-1 Hyperion Hyperspectral Imager, NASA JPL's Airborne Visible/Infrared Imaging Spectrometer, NASA's Moderate Resolution Imaging Spectrometer, and NASA's HyspIRI satellite. Two commercial hyperspectral sensors are discussed including: Itres, Inc. CASI-1500, and HyVista, Inc., HyMap, Inc. Additional information about several of these sensors is provided in Chapter 11.

NASA EO-1 Hyperion Hyperspectral Imager

NASA's Earth Observation (EO-1) satellite carried the Hyperion hyperspectral remote sensing system.

Hyperion

The Hyperion sensor provides high-quality calibrated data that can support evaluation of hyperspectral technology for Earth observing missions. Hyperion is a pushbroom instrument (Figure 2-9f) with a swath width of 7.5 km. Hyperion has a single telescope and two spectrometers, one visible/near infrared (VNIR) spectrometer and one short-wave infrared (SWIR) spectrometer. A dichroic filter in the system reflects the energy from 400 to 1,000 nm to the VNIR spectrometer and transmits the energy from 900 – 2,500 nm to the SWIR spectrometer. There are 198 useful bands out of 220 that are calibrated, including channels 8–57 for the VNIR and channels 77–224 for the SWIR (http://eo1.usgs.gov/faq). The SWIR overlap with the VNIR from 900 – 1,000 nm allows cross calibration between the two spectrometers. The spectral range of the instrument extends from 400 – 2,500 nm with a spectral resolution of 10 nm (Table 2-15). The Hg:Cd:Te detectors in the SWIR spectrometer are cooled by an advanced TRW cryocooler. The data typically are processed into cubes (19.8 km long by 7.5 km wide) to facilitate data handling in current desktop computers. Each cube consists of 75 MB of data. A typical acquisition consists of multiple cubes.

NASA Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)

The first airborne imaging spectrometer (AIS) was built to test the concept with infrared area arrays (Vane and Goetz, 1993). The spectral coverage of the instru-

TABLE 2–15 Characteristics of the NASA EO-1 Hyperion hyperspectral sensor, NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), ITRES Research, Ltd. CASI-1500, and HyVista's, Inc. HyMap.

Sensor	Technology	Spectral Resolution (nm)	Spectral Interval (nm)	Number of Bands	Quantization (bits)	IFOV (mrad)	Total field of view (°)
Hyperion	Linear array	400 – 2,500	10	220	11		
AVIRIS	Whiskbroom linear array	400 – 2,500	10	224	12	1.0	30°
CASI-1500	Linear array (1,500) and area array CCD (1,500 × 288)	370 – 1,050	2.2	288 possible; the number of bands and the number of pixels in the across-track are programmable.	14	0.49	40°
HyMap	Whiskbroom linear array	450 – 2,480	13-17	128		2.5	61.3°

ment was 1.9 – 2.1 μm in the *tree mode* and 1.2 – 2.4 μm in *rock mode* in contiguous bands that were 9.3 nm wide.

AVIRIS

To acquire data with greater spectral and spatial coverage, AVIRIS was developed at NASA's Jet Propulsion Laboratory in Pasadena, CA (Table 2-15). Using a whiskbroom scanning mirror and linear arrays of silicon (Si) and indium-antimonide (InSb) configured as in Figure 2-9e, AVIRIS acquires images in 224 bands, each 10 nm wide in the 400 – 2,500 nm region (NASA AVIRIS, 2014). The sensor is typically flown onboard the NASA/ARC ER-2 aircraft at 20 km above ground level and has a 30° total field of view and an instantaneous-field-of-view of 1.0 mrad, which yields 20 × 20 m pixels. The data are recorded in 12-bits (values from 0 to 4,095).

Moderate Resolution Imaging Spectrometer (MODIS)

The Moderate Resolution Imaging Spectrometer (MODIS) is flown on NASA's EOS *Terra* (a.m. equatorial crossing time) and *Aqua* (p.m. equatorial crossing time) satellites (Table 2-15). MODIS provides long-term observations to obtain knowledge about global dynamics and processes occurring on the surface of the Earth and in the lower atmosphere (NASA MODIS, 2014). MODIS yields information on atmospheric (cloud cover and associated properties), oceanic (sea-surface temperature and chlorophyll), and land-surface (land-cover changes, land-surface temperature, and vegetation properties) characteristics.

MODIS is in a 705-km Sun-synchronous orbit. It views the entire surface of the Earth every 1 to 2 days. It has a field of view of ±55° off-nadir, which yields a swath width of 2,330 km (Figure 2-50a). MODIS obtains high radiometric resolution images (12-bit) of daylight-reflected solar radiation and day/night thermal emission over all regions of the globe. MODIS is a whiskbroom scanning imaging radiometer consisting of a cross-track scan mirror, collecting optics, and a set of linear detector arrays with spectral interference filters located in four focal planes. MODIS collects data in 36 co-registered spectral bands: 20 bands from 0.4 – 3 μm and 16 bands from 3 – 15 μm. The spatial and spectral resolution of the bands and their primary uses are summarized in Table 2-16.

MODIS' coarse spatial resolution ranges from 250 × 250 m (bands 1 and 2) to 500 × 500 m (bands 3 through 7) and 1 × 1 km (bands 8 through 36). A MODIS image of the Nile Delta, the Nile River, and much of the area surrounding the Red Sea is shown in Figure 2-50b. MODIS provides daylight reflection and day/night emission spectral imaging of any point on Earth at least every 2 days, with a continuous duty cycle.

MODIS has one of the most comprehensive calibration subsystems ever flown on a satellite remote sensing system. The calibration hardware includes a solar diffuser, a solar diffuser stability monitor, a spectroradiometric calibration instrument, a black body for thermal calibration, and a space viewport. The calibration allows the optical MODIS data to be converted into scaled percent reflectance. MODIS data are being

TABLE 2–16 Characteristics of the *Terra* satellite Moderate Resolution Imaging Spectrometer (MODIS).

Band	Spectral Resolution (μm)	Spatial Resolution	Band Utility
1 2	0.620 – 0.670 0.841 – 0.876	250 × 250 m 250 × 250 m	Land-cover classification, chlorophyll absorption, leaf-area-index mapping
3 4 5 6 7	0.459 – 0.479 0.545 – 0.565 1.230 – 1.250 1.628 – 1.652 2.105 – 2.155	500 × 500 m 500 × 500 m 500 × 500 m 500 × 500 m 500 × 500 m	Land, cloud, and aerosol properties
8 9 10 11 12 13 14 15 16	0.405 – 0.420 0.438 – 0.448 0.483 – 0.493 0.526 – 0.536 0.546 – 0.556 0.662 – 0.672 0.673 – 0.683 0.743 – 0.753 0.862 – 0.877	1 × 1 km 1 × 1 km 1 × 1 km 1 × 1 km 1 × 1 km 1 × 1 km 1 × 1 km 1 × 1 km 1 × 1 km	Ocean color, phytoplankton, biogeochemistry
17 18 19	0.890 – 0.920 0.931 – 0.941 0.915 – 0.965	1 × 1 km 1 × 1 km 1 × 1 km	Atmospheric water vapor
20 21 22 23	3.600 – 3.840 3.929 – 3.989 3.929 – 3.989 4.020 – 4.080	1 × 1 km 1 × 1 km 1 × 1 km 1 × 1 km	Surface–cloud temperature
24 25	4.433 – 4.498 4.482 – 4.549	1 × 1 km 1 × 1 km	Atmospheric temperature
26	1.360 – 1.390	1 × 1 km	Cirrus clouds
27 28 29	6.535 – 6.895 7.175 – 7.475 8.400 – 8.700	1 × 1 km 1 × 1 km 1 × 1 km	Water vapor
30	9.580 – 9.880	1 × 1 km	Ozone
31 32	10.780 – 11.280 11.770 – 12.270	1 × 1 km 1 × 1 km	Surface–cloud temperature
33 34 35 36	13.185 – 13.485 13.485 – 13.785 13.785 – 14.085 14.085 – 14.385	1 × 1 km 1 × 1 km 1 × 1 km 1 × 1 km	Cloud-top altitude

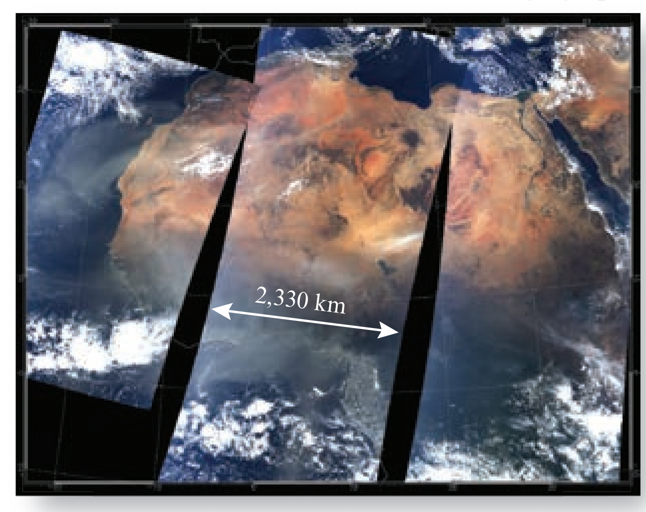
processed to create numerous global datasets, including (NASA MODIS, 2014):

- land ecosystem variables (e.g., vegetation indices, LAI, fraction of photosynthetically active radiation, vegetation net primary production),
- atmospheric variables (e.g., cloud fraction, cloud optical thickness, aerosol optical depth, etc.), and
- ocean variables (e.g., sea-surface temperature and chlorophyll).

NASA Hyperspectral Infrared Imager (HyspIRI)

NASA's Hyperspectral Infrared Imager (HyspIRI) is to be launched in the future. It is designed to study the world's ecosystems and provide information on natural disasters, such as volcanoes, wildfires and drought, and vegetation health. The sensor was recommended in the National Research Council Decadal Survey requested by NASA, NOAA, and USGS. The HyspIRI mission includes two instruments mounted on a satellite in low

Moderate Resolution Imaging Spectrometer (MODIS) Imagery




a. Three swaths of MODIS imagery over northern Africa on February 29, 2000.



b. MODIS image of the Nile Delta obtained on February 28, 2000.

FIGURE 2–50 a) A MODIS composite image for the North African Continent on February 29, 2000. A dust storm can be seen blowing off the coast of Morocco in the northwest corner. The swath width is 2,330 km. b) MODIS image of the Nile Delta on February 28, 2000 (images courtesy of NASA Visible Earth).

Earth orbit (NASA HyspIRI, 2014). There is an imaging spectrometer measuring from the visible to short-wave infrared (VSWIR: 380 nm – 2,500 nm) in 10 nm contiguous bands and a multispectral imager measuring from 3 – 12 μm in the mid and thermal infrared (TIR). The TIR sensor will have eight bands. Both VSWIR and TIR instruments have a spatial resolution of 60 m at nadir. The VSWIR will have a revisit of 19 days and the TIR will have a revisit of 5 days. HyspIRI also includes an Intelligent Payload Module (IPM) which will enable direct broadcast of a subset of the data (Glavich et al., 2009).

Itres, Inc. Compact Airborne Spectrographic Imager-1500

ITRES Research, Ltd., of Canada has developed an innovative variety of innovative visible near-infrared (CASI-1500), short-wavelength infrared (SASI-600), mid-wavelength infrared (MASI-600), and thermal infrared (TASI-600) hyperspectral instruments.

CASI-1500

CASI-1500 is a pushbroom imaging spectrometer based on the use of a 1,500 × 288 area array CCD. The instrument operates over a 670-nm spectral range (380 – 1,050 nm) and has a 40° total field of view across 1,500 pixels (Itres CASI-1500, 2014). A single line of terrain 1,500 pixels wide perpendicular to the flight path is sensed by the spectrometer optics (Figure 2-51). The incoming radiant flux from each pixel is then spectrally dispersed along the axis of the area array CCD (Figure 2-51) so that a spectrum of energy (from blue through near-infrared) is obtained for each pixel across the swath. By repetitively reading the contents of the area array CCD as the aircraft moves along the flight path, a two-dimensional image at high spectral resolution is acquired. Because the radiant flux for all pixels in a particular swath are recorded simultaneously, spatial and spectral co-registration is assured. The across-track spatial resolution is determined by the altitude of the CASI above ground level and the IFOV, while the along-track resolution depends upon the velocity of the aircraft and the rate at which the CCD is read.

CASI-1500 is spectrally programmable meaning that the user can specify which bands are to be collected for a specific application (e.g., vegetation biomass monitoring, bathymetric mapping, inventorying chlorophyll *a* concentration, urban/suburban analysis). The result is a programmable area array VNIR hyperspectral remote sensing system.

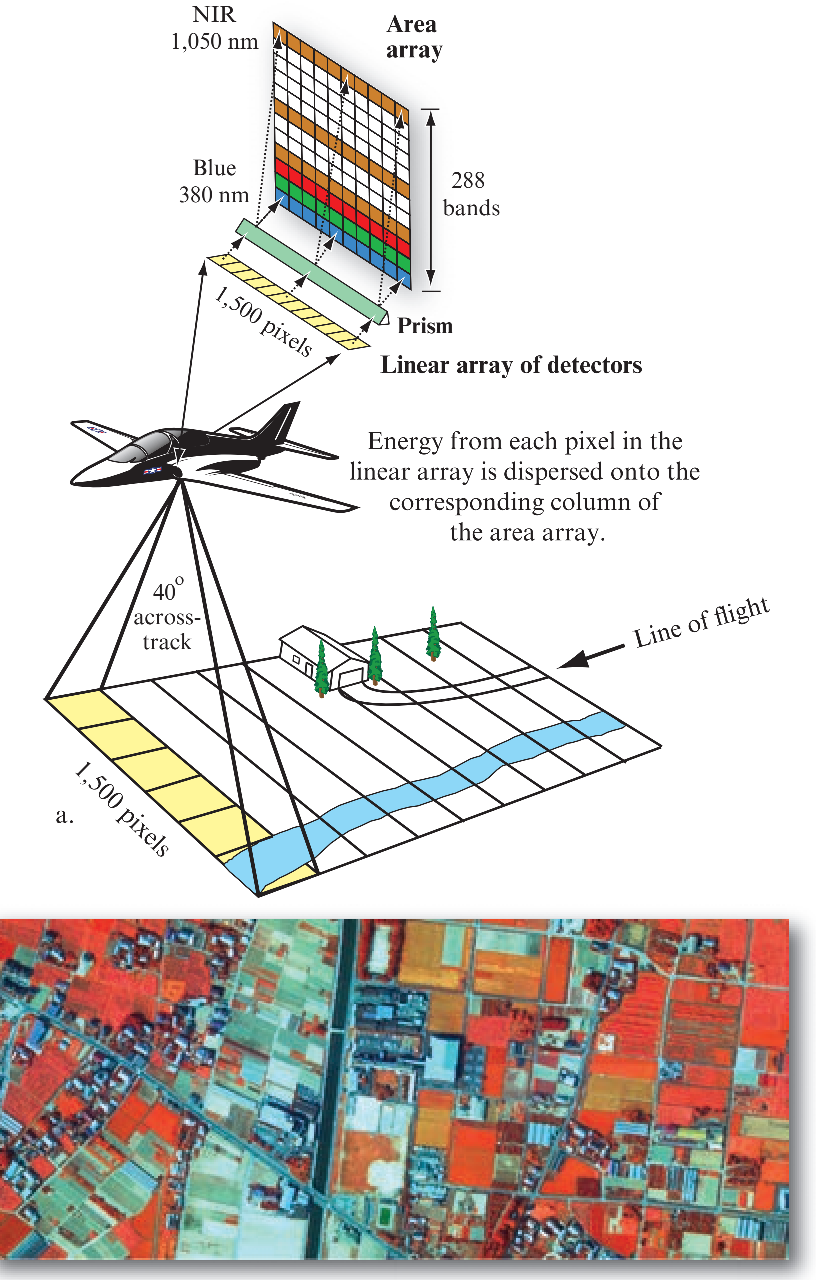
SASI-600

SASI-600 collects 100 bands of SWIR hyperspectral data from 950 – 2,450 nm at 15-nm intervals with 600 across-track pixels. SASI-600 data are especially useful for geological exploration and plant speciation (Itres SASI-600, 2014).

MASI-600

MASI-600 is the first commercially available mid-wavelength hyperspectral sensor designed specifically

Hyperspectral Data Collection Using Linear and Area Arrays
ITRES Research Ltd., Calibrated Airborne Spectrographic Imager (CASI) 1500



b. CASI-1500 false-color composite of an area in Japan.

FIGURE 2–51 a) Hyperspectral data collection using a linear array pushbroom sensor that distributes the energy onto a 1,500 × 288 element area array sensitive to the wavelength region from 380 to 1,050 nm. b) CASI-1500 image of an area in Japan (image courtesy of ITRES Research, Ltd.).

for airborne use. It records 64 bands in the 4 – 5 μm range with a 40° across-track field of view encompassing 600 pixels (Itres MASI-600, 2014).

TASI-600

The TASI-600 is a pushbroom hyperspectral thermal sensor system designed specifically for airborne use. It collects 32 or 64 bands in the region from 8 – 11.5 μm with a 40° across-track field of view encompassing 600 pixels (Itres TASI-600, 2014).

HyVista, Inc., HyMap

HyVista Inc. of Australia operates a HyMap whiskbroom hyperspectral scanner (Figure 2-52) manufactured by Integrated Spectronics Pty Ltd. (HyVista

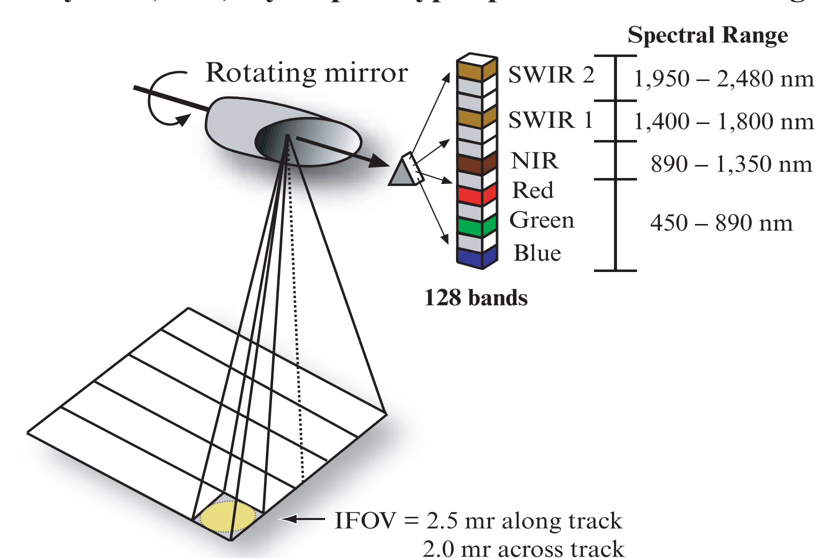


FIGURE 2–52 HyMap hyperspectral remote sensing using whiskbroom technology in the region from 450 – 2,480 nm.

HyMap, 2014). The HyMap sensor records 128 bands in the reflective region from 450 – 2,480 nm with contiguous spectral coverage (except in the atmospheric water vapor bands near 1,400 and 1,900 nm) with average spectral sampling intervals between 13 – 17 nm. Typical ground spatial resolution is between 2 and 10 m depending upon flight altitude AGL.



Airborne Digital Cameras

Digital cameras have revolutionized airborne photogrammetric data collection. The imaging sensor of a digital camera is usually a charge-coupled device (CCD) or Complementary Metal-Oxide-Semiconductor (CMOS) computer chip. The sensor detectors convert light into electrons that can be measured and converted into a radiometric intensity value. Digital cameras may be based on linear or area arrays of detectors (Figure 2-9b).

Digital cameras utilize a lens with its associated diaphragm to control the *f*/stop, a shutter to control the length of exposure, and a focusing device. However, the major difference is that instead of using film, CCD linear or area arrays are located at the film plane. The lens focuses the light from the outside world onto the linear- or area-array bank of detectors. The photons of light illuminating each of the detectors cause an electrical charge to be produced that is directly related to the amount of incident radiant energy. This analog signal is then sampled electronically and converted into a digital brightness value ranging from 8-bit (values from 0 – 255) to 12-bit (values from 0 to 4,095). The brightness values obtained from the analog-to-digital (A-to-D) conversion may be stored and read by computer systems. The CCDs are more sensitive to spectral reflectance changes in the scene than the silver halide crystals used in conventional analog color and color-infrared aerial photography.

Small-Format Digital Cameras

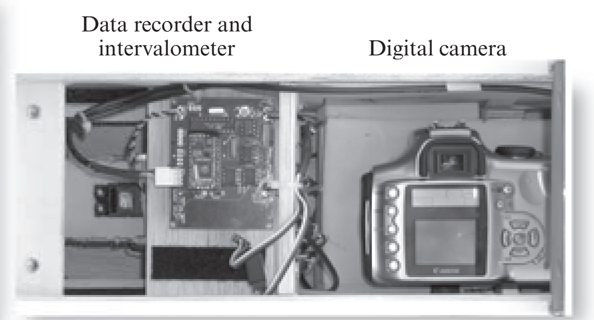
These cameras typically collect digital images that are < 16 megapixels (MP) per band. For example, the Canon digital camera shown in Figure 2-53b is based on an area array with 3,456 columns and 2,304 rows yielding approximately 8 MP per band. At the instant of exposure, the camera rapidly records three versions of the scene using internal filtration. The result is one image based solely on blue light reflected from the terrain, another based on only green light reflected from the terrain, and a final image produced only from reflected red light. The three individual black-and-white images are recorded in the camera's random-access memory (RAM) and can be color-composited using additive color theory to produce a natural-looking color photograph. It is also possible to make the detectors sensitive to near-infrared light.

The inexpensive unmanned aircraft system (UAS) shown in Figure 2-53ab contains an intervalometer to obtain photographs at specific intervals to obtain the necessary end-lap. The images are stored on a data recorder. The operator on the ground knows where the

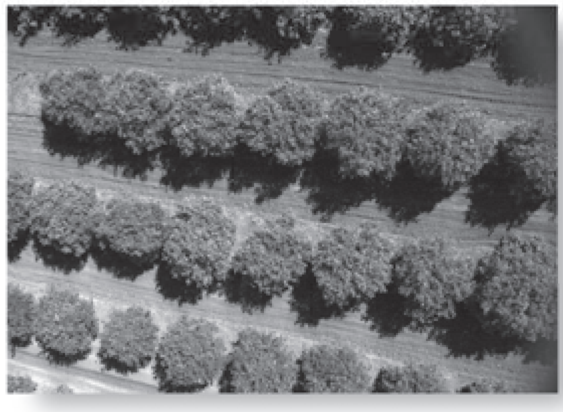
Inexpensive Unmanned Aerial System (UAS) using a Small-Format Digital Frame Camera



a. Unmanned aerial system fuselage.



b. System components.



c. Citrus grove near Weslaco, TX.



d. Water treatment plant on South Padre Island, TX.

FIGURE 2–53 a) An unmanned aerial system (UAS) modified to carry a small-format digital camera. b) System components include a data recorder, intervalometer, and GPS antenna (not shown). cd) Examples of aerial photography obtained using the unmanned aerial system (courtesy of Perry Hardin and Mark Jackson, Department of Geography, Brigham Young University).

plane is based on the use of an onboard GPS antenna (not shown). The inexpensive digital camera obtains high-resolution aerial imagery because it is usually flown relatively close to the ground (Figures 2-53cd).

Medium-Format Digital Cameras

These cameras are usually based on linear or area arrays with greater than 4,000 × 4,000 detectors per band (yielding 16 MP).

Leica Geosystems, Ag., RCD30

Leica Geosystems, Ag., produces several of the most popular medium- and large-format remote sensing digital data-collection sensor systems used by many photogrammetric engineering firms in the United States (Leica, 2014a). For example, much of the imagery acquired for public use (e.g., USDA National Agriculture Imagery Program [NAIP]), counties, states) is collected using Leica digital cameras.The RCD30 provides multispectral data in four co-registered bands (RGB and NIR) with each digital frame of imagery being 8,956 × 6,708 pixels in dimension. The camera has forward motion compensation (FMC) along two axes (forward and lateral). It is one of the most accurate and most reliable medium-format cameras available (Leica RCD30, 2014).

Large-Format Digital Cameras

Large-format digital cameras typically have very large CCD linear or area CCD arrays. These cameras surpass the performance of large-format analog frame cameras. Because of the large number of large-format analog cameras already in existence, digital cameras will be used along with the traditional analog cameras for some time. Eventually, however, most photogrammetric aerial photography will be obtained using large-format digital metric cameras. Two companies that produce some of the most important medium- to large-scale digital cameras are Leica Geosystems, Ag, and Microsoft Ultramap, Inc.

Leica Geosystems, Ag., ADS80, Z/I Imaging DMC Aerial Photography

The Leica Airborne Digital Sensor—ADS80 uses linear array technology to collect data in panchromatic, red, green, blue (RGB) and near-infrared (NIR) bands.

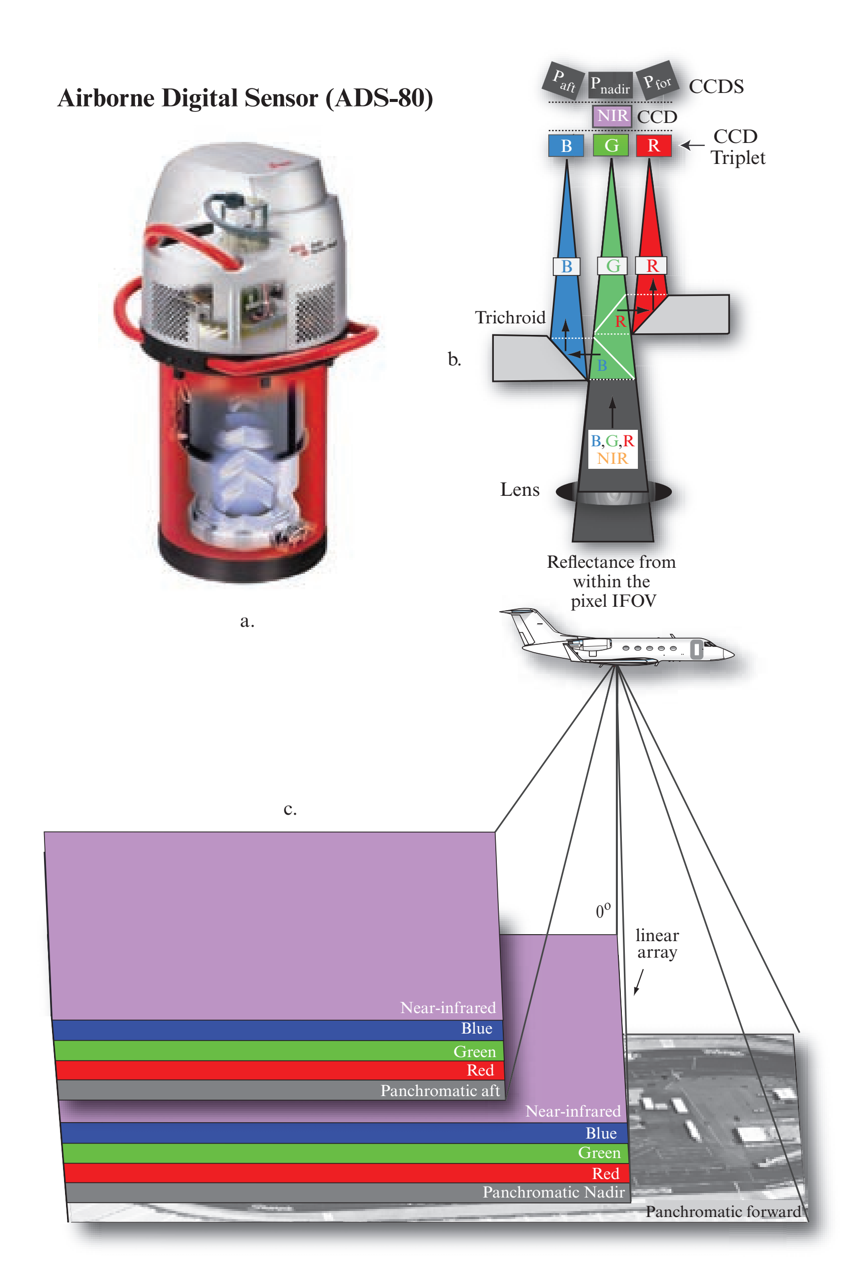
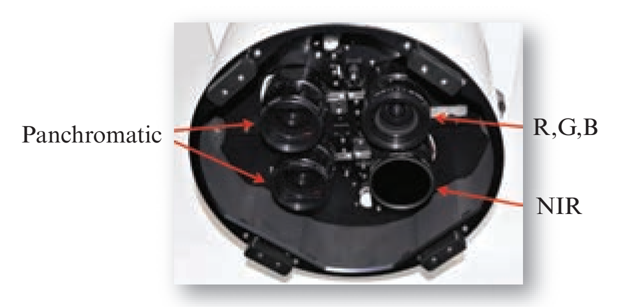


FIGURE 2–54 a) Leica Geosystems Airborne Digital Sensor—ADS-80. b) Dispersion of incoming panchromatic, blue, green, red, and near-infrared energy onto the 12,000 detectors of each of the linear arrays. c) Orientation of the aft, Nadir, and forward-looking linear arrays (images courtesy of Leica Geosystems).

Each of the CCD linear arrays consists of 12,000 pixels. Each of the pixels in the linear arrays is 6.5 μm in size (Leica, 2014a). The data are collected using the nadir, aft- and forward-looking logic shown in Figure 2-54. All of the panchromatic and spectral data are in near-perfect registration at the same radiometric resolution (Leica ADS80, 2014)

The Z/1 Imaging Digital Mapping Camera (DMC) family includes four different configurations that collect metric digital frame camera imagery. The base Z/1 RMK D (5,760 × 6,400 pixels) is a four-camera-head multispectral sensor that collects RGB and NIR imagery. The Z/I DMC II140 consists of a five-camera-head multispectral sensor (RGB, NIR) (each 6,846 × 6,096 pixels) plus a 140 megapixel panchromatic sensor (12,096 × 11,200 pixels) that doubles the ground resolution of the Z/I RMK D. The Z/1 DMC II230 is a five-camera-head multispectral sensor (RGB, NIR) (each 6,846 × 6,096 pixels) that has a 230 megapixel CCD (15,552 × 14,144 pixels) that increases ground coverage even more to reduce flight lines. The Z/1 DMC II250 is a five-camera-head multispectral sensor (RGB, NIR) (each 6,846 × 6,096 pixels) that has a panchromatic band (16,768 × 14,016 pixels). It has longer

UltraCam™ Digital Frame Camera
Used to Obtain Bing Maps Aerial Photography



a. UltraCam panchromatic and multispectral lenses.



b. UltraCam natural color digital photography. c. UltraCam color-infrared digital photography.

FIGURE 2–55 Characteristics of the Microsoft, Inc. UltraCam™ digital frame camera. a) Lenses associated with panchromatic, color, and color-infrared data collection. b–c) Examples of natural color and color-infrared UltraCam™ aerial photography (images courtesy of Microsoft Photogrammetry Division, www.iflyultracam.com; updated from Jensen et al., 2012).

focal length optics, yielding even greater coverage with finer detail at lower flight altitudes (Leica Z/1, 2014).

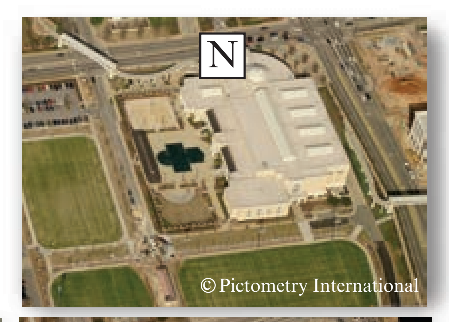
Microsoft, Inc., UltraCam Eagle

The Microsoft, Inc. Photogrammetry Division developed the UltraCam digital frame camera primarily to obtain imagery for its *Bing Maps* platform search engine. In addition, the division markets UltraCam cameras and related products to the photogrammetric engineering industry, including: UltraCam Eagle (the newest sensor), ULTRACAM Xp, ULTRACAM Xp-WA, and ULTRACAM Lp (Microsoft UltraCam, 2014).

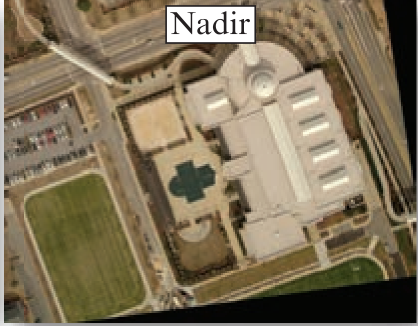
UltraCam imagery is the data source for automated 3D framework production in *Bing Maps* (Wiechert, 2009a). They use 80% endlap and 60% sidelap UltraCam photography (yielding up to 12 rays per ground pixel based on multi ray photogrammetry) to extract extremely high-resolution digital surface models (DSM) of the terrain with >50 points per m^2 (Wiechert 2009b). Derived DSM and digital terrain models (DTM) in conjunction with image classification (consisting of 10–15 classes such as buildings, vegetation, concrete, etc.) are used to create the three-dimensional roof polygons, extrude completed roof geometry to the DTM, and apply photo texture from UltraCam images to create a three-dimensional Virtual Earth served in *Bing Maps* (Wiechert, 2009ab). UltraCam photogrammetry is performed using the ULTRAMAP Workflow Software System.

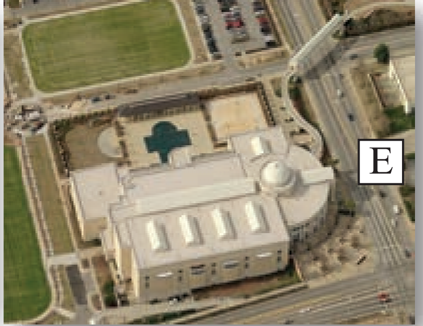
The UltraCam Eagle system has one of the largest panchromatic image footprints when compared with

Pictometry Digital Aerial Photography of Columbia, SC











Photography was obtained from multiple flightlines.

FIGURE 2–56 Natural color vertical and oblique aerial photography of the Strom Thurmond Wellness Center in Columbia, SC, obtained at four cardinal directions (North, East, South, West) and at Nadir (i.e., vertical). The nominal spatial resolution was 6 × 6 in. (photography courtesy of Pictometry International, Inc.).

the other ultra-large format metric camera with 20,010 × 13,080 detectors. This can reduce the number of data-collection flight lines required for a given project. Multispectral data are collected in the blue, green, red, and reflective near-infrared regions of the spectrum using sensors with 6,67 × 4,360 detectors. Radiometric resolution is 8- to 16-bits (Microsoft UltraCam, 2014). An example of the quality of UltraCam Eagle multispectral digital frame camera photography is shown in Figure 2-55. The frame data-collection rate is 1.8 seconds which allows 80% to 90% forward overlap if desired. Microsoft markets other digital frame cameras (not discussed here) (Microsoft UltraCam, 2014).

Digital Oblique Aerial Photography

Sometimes people feel more comfortable looking at and analyzing oblique aerial photography rather than vertical aerial photography (Figure 2-56). Several commercial vendors now collect vertical as well as oblique aerial photography.

Pictometry International, Inc. Oblique and Vertical Aerial Photography

Pictometry uses a five-camera arrangement whereby every time a vertical aerial photograph is collected, four additional oblique photographs are obtained in the North, East, South, and West directions (Pictometry, 2014). The flight lines overlap by 20 to 30%, which allows each feature in the landscape such as a building to be recorded and viewed from many different vantage points. The image analyst can simply select the view that is most pleasing and provides the most useful thematic information. Pictometry aerial photography of a portion of Columbia, SC, is shown in Figure 2-56. Innovations in digital image processing make it possible to register the oblique aerial photography in a GIS where both horizontal and vertical measurements of

Digital Frame Camera Image of the Toquepala Copper Mine in Southern Peru Taken from the International Space Station

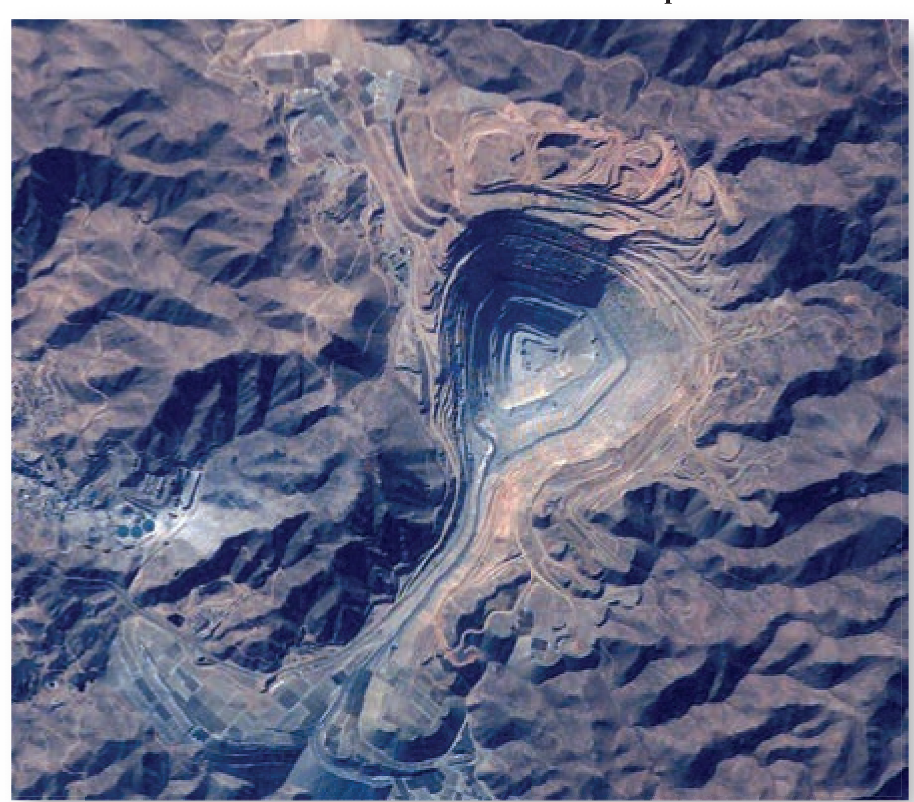
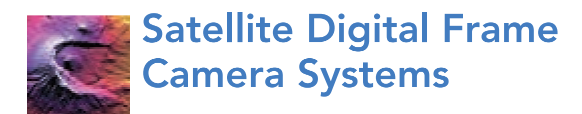


FIGURE 2–57 Astronaut photograph of the Toquepala Copper Mine in Southern Peru taken from the International Space Station on September 22, 2003, with a Kodak DCS 760 digital camera (photo #ISS007-E-15222). The open pit mine is 6.5 km across and descends more than 3,000 m into the earth. A dark line on the wall of the pit is the main access road to the bottom. Spoil dumps of material mined from the pit are arranged in tiers along the northwest lip of the pit (courtesy of NASA Earth Observatory and Dr. Kamlesh Lulla, NASA Johnson Space Center).

structures and the terrain can be made while viewing the oblique photography.



Satellite Digital Frame Camera Systems

Despite the ongoing development of electro-optical remote sensing instruments, traditional optical camera systems continued to be used for space-survey purposes. For example, the U.S. Space Shuttle and Space Station astronauts routinely collected photography using Hasselblad and Linhof cameras.

U.S. Space Shuttle Photography

NASA astronauts routinely documented Earth processes during Space Transportation System (STS) missions using both analog and digital camera systems. These efforts resulted in an impressive database of more than 400,000 Earth images. Photographic documentation of Earth processes during manned spaceflights was the cornerstone of the Space Shuttle Observations program, as it was with the earlier *Mercury*, *Gemini*, *Apollo*, and *Skylab* Earth observations programs (Lulla and Dessinov, 2000). During the space shuttle era, more than 250 selected sites of interest to geoscientists were identified. Data from these sites were acquired during Space Shuttle missions and cataloged into a publicly accessible electronic database according to the specific mission (e.g., STS-74) or by thematic topic (NASA Shuttle Photography, 2014).

Space Shuttle Analog Cameras

The primary analog cameras used during Space Shuttle missions were the Hasselblad and Linhof systems. NASA-modified Hasselblad 500 EL/M 70-mm cameras were used with large film magazines, holding 100 to 130 exposures.

The four windows in the aft part of the Space Shuttle were used to obtain photography of the Earth. The windows allowed only 0.4 to 0.8 μm light to pass

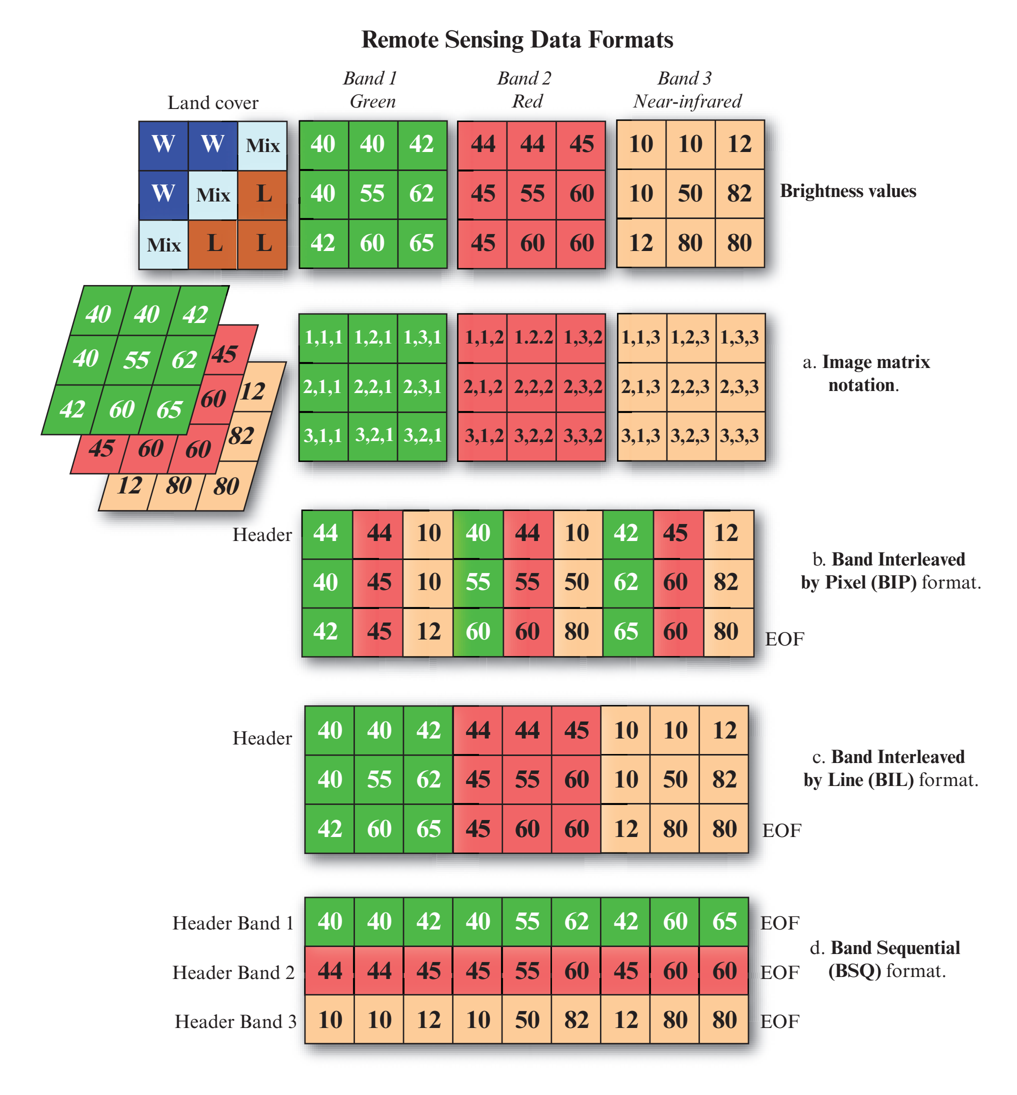


FIGURE 2–58 Selected digital image data formats.

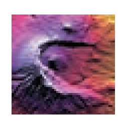
through. This resulted in the use of two primary film bases in the Hasselblad and Aero-Technika Linhof camera systems, including visible color (Kodak 5017/6017 Professional Ektachrome) and color-infrared (Kodak Aerochrome 2443) films.

Space Shuttle photographs were obtained at a variety of Sun angles, ranging from 1° to 80°, with the majority of pictures having Sun angles of approximately 30°. Very low Sun angle photography often provides unique topographic views of remote mountainous areas otherwise poorly mapped. Sequential photographs with different look angles can provide stereoscopic coverage. Seventy-five percent of the photographs in the archive cover the regions between 28°N and 28°S latitude, providing coverage for many little-known tropical areas. Twenty-five percent of the images cover regions between 30° to 60° N and S latitude.

The Space Shuttle Earth Observations Project (SSEOP) photography database of the Earth Science Branch at the NASA Johnson Space Center contains the records of more than 400,000 photographs of the Earth made from space. A select set of photographs has been digitized and placed in the public domain (http://images.jsc.nasa.gov).

Space Shuttle and Space Station Digital Photography

The International Space Station (ISS) was launched November 2, 2000. It continues the NASA tradition of Earth observation from human-tended spacecraft. The ISS U.S. Laboratory Module has a specially designed optical window with a clear aperture 50.8 cm in diameter that is perpendicular to the Earth's surface most of the time. In 2001, Space Shuttle astronauts began acquiring digital images that were transmitted directly to the ground while in orbit. An astronaut photograph of the Toquepala Copper Mine in Southern Peru taken from the International Space Station on September 22, 2003, with a Kodak DCS 760 digital camera is shown in Figure 2-57.



Digital Image Data Formats

The image analyst can order digital remote sensor data in a variety of formats. The most common formats are (Figure 2-58):

- Band Interleaved by Pixel (BIP),
- Band Interleaved by Line (BIL), and
- Band Sequential (BSQ).

To appreciate the data formats, consider a hypothetical remote sensing dataset containing just nine pixels obtained over the interface between land and water (Figure 2-58). The image consists of three bands (band 1 = green; band 2 = red; band 3 = near-infrared). The brightness value ($BV_{i,j,k}$) row, column, and band notation is provided (Figure 2-58a).

Band Interleaved by Pixel Format

The BIP format places the brightness values in *n* bands associated with each pixel in the dataset in sequential order For example, if a dataset contained three bands the format for the first pixel in the matrix (1,1) is 1,1,1; 1,1,2; 1,1,3. The brightness values for pixel (1,2) are then placed in the dataset (e.g., 1,2,1; 1,2,2; and 1,2,3), and so on. An end-of-file (EOF) marker is placed at the end of the dataset (Figure 2-58b).

Band Interleaved by Line Format

The BIL format creates a file that places the brightness values in *n* bands associated with each line in the dataset in sequential order. For example, if there are three bands in the dataset, all of the pixels in line 1, band 1 are followed by all of the pixels in line 1, band 2, and then line 1, band 3. An EOF marker is placed at the end of the dataset (Figure 2-58c).

Band Sequential Format

The BSQ format places all of the individual pixel values in each band in a separate and unique file. Each band has its own beginning header record and EOF marker (Figure 2-58d).



Summary

The public and scientists should perform digital image processing of remote sensor data only after they understand and appreciate exactly how the imagery was collected. This chapter provided information about the characteristics of many important digital remote sensing systems and the most often used digital data formats.



References

Astrium, 2014, *Ecotechnology and Satellite Imagery*, Germany: Airbus Defense and Space, http://www.astrium-geo.com/.

Astrium Pleiades, 2014, *Pleiades Satellite Imagery*, Germany: Airbus Defense and Space, www.astrium-geo.com/pleiades/.

Astrium SPOT, 2014, *SPOT Satellite Imagery*, Germany: Airbus Defense and Space, www.astrium-geo.com/en/143-spot-satellite-imagery.

Astrium SPOT 6&7, 2014, *SPOT 6 and Spot 7 Imagery*, Germany: Airbus Defense and Space, www.astrium-geo.com/en/147-spot-6-7-satellite-imagery.

CNES Pleiades, 2014, *Pleiades: Dual Optical System for Metric Resolution Observations*, Paris: Centre National D'Etudes Spatiales, http://smsc.cnes.fr/PLEIADES/index.htm.

Digenis, C. J., 2005, "The EO-1 Mission and the Advanced Land Imager," *Lincoln Journal*, 15(2):161–164.

DigitalGlobe QuickBird, 2014, *QuickBird Specifications*, www.digitalglobe.com/sites/default/files/Basic%20Imagery%20Datasheet_0.pdf.

DigitalGlobe WorldView-1, 2014, *WorldView-1 Specifications*, www.digitalglobe.com/sites/default/files/Basic Imagery-DS-BASIC-PROD.pdf.

DigitalGlobe WorldView-2, 2014, *WorldView-2 Specifications*, www.digitalglobe.com/sites/default/files/Basic Imagery-DS-BASIC-PROD.pdf.

DigitalGlobe WorldView-3, 2014, *WorldView-3 Specifications*, www.digitalglobe.com/about-us/content-collection#worldview-3.

DMC, 2014a, *DMC SLIM-6*, Guildford, Surrey, UK: DMC International Imaging, www.dmcii.com.

DMC, 2014b, *NigeriaSat-2 Sensor & Data Characteristics*, Guildford, Surrey, UK: DMC International Imaging, www.dmcii.com.

Dykstra, J., 2012, "Comparison of Landsat and RapidEye Data for Change Monitoring Applications" presented at the *11th Annual Joint Agency Commercial Imagery Evaluation (JACIE) Workshop*, Fairfax, VA, April 17–19, 32 p.

eoPortal KOMPSAT-5, 2015, *KOMPSAT-5,* https://directory.eoportal.org/web/eoportal/satellite-missions/k/kompsat-5.

ESA CEOS, 2014, *Earth Observation Handbook*, www.eohandbook.com.

ESA GMES Sentinel-2, 2014, *GMES Sentinel-2 Mission Requirements Document*, European Space Agency, http://esa-multimedia.esa.int/docs/GMES/Sentinel-2_MRD.pdf.

ESA Sentinel-2, 2014, *Sentinel-2*, European Space Agency, https://earth.esa.int/web/guest/missions/esa-future-missions/sentinel-2.

Gasparovic, R. F., 2003, *Ocean Remote Sensing Program*, Johns Hopkins University Applied Physics Laboratory, http://fermi.jhuapl.edu/avhrr/index.html.

GeoEye, 2014, *GeoEye*, www.digitalglobe.com/sites/default/files/Standard%20Imagery%20Datasheet_0.pdf.

Glavich, T., Green, R. O., Hook, S. J., Middleton, B., Rogez, F., and S. Ungar, 2009, *HyspIRI Decadal Survey Mission Development Status*, Washington: Decadal Survey Symposium, February 11–12, 39 p.

Gleason, J., Butler, J., and N. C. Hsu, 2010, *NPP/VIIRS: Status and Expected Science Capability*, Greenbelt: NASA Goddard Space Flight Center, http://modis.gsfc.nasa.gov/sci_team/meetings/201001/presentations/plenary/gleason.pdf.

HyVista HyMAP, 2014, *HyMap*, Sydney: HyVista, Inc., www.hyvista.com.

ImageSat EROS, 2014, *ImageSat International: EROS A and B*, www.imagesatintl.com.

Irons, J. R., Dwyer, J. L., and J. A. Barsi, 2012, "The Next Landsat satellite: The Landsat Data Continuity Mission," *Remote Sensing of the Environment*, 122(2012):11–21.

ISRO IRS-1C, 2014, *IRS-1C*, India: National Remote Sensing Agency, http://www.isro.org/satellites/irs-1c.aspx.

ISRO IRS-1D, 2014, *IRS-1D,* India: National Remote Sensing Agency, http://www.isro.org/satellites/irs-1d.aspx.

ISRO CARTOSAT-1, 2014, *CARTOSAT-1*, India: National Remote Sensing Agency, http://www.isro.org/satellites/earthobservationsatellites.aspx.

ISRO CARTOSAT-2B, 2014, *CARTOSAT-2B*, Bangalore, India: Indian Space Research Organization, www.isro.gov.in/satellites/earthobservationsatellites.aspx.

ISRO RESOURCESAT-1, 2014, *RESOURCESAT-1*, India: National Remote Sensing Agency, http://www.isro.org/satellites/earthobservationsatellites.aspx.

ISRO RESOURCESAT-2, 2014, *RESOURCESAT-2*, India: National Remote Sensing Agency, http://www.isro.org/satellites/earthobservationsatellites.aspx.

ITRES CASI-1500, 2014, *CASI-1500*, Canada: ITRES Research Ltd., www.itres.com.

ITRES MASI-600, 2014, *MASI-600*, Canada: ITRES Research Ltd., www.itres.com.

ITRES SASI-600, 2014, *SASI-600*, Canada: ITRES Research Ltd., www.itres.com.

ITRES TASI-600, 2014, *TASI-600*, Canada: ITRES Research Ltd., www.itres.com.

Jensen, J. R. and R. R. Jensen, 2013, *Introductory Geographic Information Systems*, Boston: Pearson, 400 p.

Jensen, J. R., Guptill, S. and D. Cowen, 2012, *Change Detection Technology Evaluation*, Bethesda: U.S. Bureau of the Census, Task 2007, FY2012 Report, 232 pages without appendices.

Kodak, 2001, *Press Release: NASA to Launch Kodak Professional DCS 760 Digital Camera on Mission to International Space Station*, www.kodak.com/US/en/corp/pressReleases/pr200106 26-01.shtml.

Leica ADS80, 2014, *ADS80: Airborne Data Sensor,* Heerbrugg, Switzerland, Leica Geosystems Ag, www.leica-geosystems.us/en/Airborne-Imaging_86816.htm.

Leica RCD30, 2014, *Leica RCD30 Series*, Heerbrugg, Switzerland, Leica Geosystems Ag, www.leica-geosystems.us/en/Airborne-Imaging_86816.htm.

Leica Z/1, 2014, *Z/1 Imaging DMC*, Heerbrugg, Switzerland, Leica Geosystems Ag, www.ziimaging.com/en/zi-dmc-iie-camera-series_20.htm.

Light, D. L., 1993, "The National Aerial Photography Program as a Geographic Information System Resource," *Photogrammetric Engineering & Remote Sensing*, 48(1):61–65.

Light, D. L., 1996, "Film Cameras or Digital Sensors? The Challenge Ahead for Aerial Imaging," *Photogrammetric Engineering & Remote Sensing*, 62(3):285–291.

Light, D. L., 2001, "An Airborne Direct Digital Imaging System," *Photogrammetric Engineering & Remote Sensing*, 67(11):1299–1305.

Loveland, T. R. and J. W. Dwyer, 2012, "Landsat: Building A Strong Future," *Remote Sensing of Environment*, 122(2012):22–29.

Lulla, K. and L. Dessinov, 2000, *Dynamic Earth Environments: Remote Sensing Observations from Shuttle-Mir Missions*, New York: John Wiley & Sons, 268 p.

Markham, B., 2011, *Landsat Data Continuity Mission Overview*, Greenbelt: GSFC, http://calval.cr.usgs.gov/JACIE_files/JACIE11/Presentations/TuePM 310_Markham_JACIE_11.080.pdf

Mendenhall, J. A., Bruce, C. F., Digenis, C. J., Hearn, D. R., and D. E. Lencioni, 2012, *Advanced Land Imagery Summary*, Lexington: MIT Lincoln Lab, http://eo1.gsfc.nasa.gov/new/validationreport/Technology/Documents/Summaries/01-ALI_Rev_0.pdf.

Microsoft UltraCam, 2014, *UltraCam* Eagle, Graz Austria: Microsoft Photogrammetry Division, www.iflyultracam.com.

NASA ASTER, 2014a, *ASTER Home Page*, Washington: NASA, http://asterweb.jpl.nasa.gov/.

NASA ASTER, 2014b, *Requesting New ASTER Acquisitions*, Washington: NASA, http://asterweb.jpl.nasa.gov/NewReq.asp.

NASA AVIRIS, 2014, *Airborne Visible/Infrared Imaging Spectrometer - AVIRIS*, Pasadena: NASA Jet Propulsion Laboratory, http://asterweb.jpl.nasa.gov/NewReq.asp.

NASA EO-1, 2014, *Earth Observing-1*, Washington: NASA, http://eo1.gsfc.nasa.gov/

NASA HyspIRI, 2014, *Hyperspectral Infrared Imagery (HyspIRI) Mission Study*, Washington: NASA, http://hyspiri.jpl.nasa.gov/.

NASA Landsat 7, 2014, *Landsat 7*, http://landsat.gsfc.nasa.gov/.

NASA MISR, 2014, *Multiangle Imaging Spectrometer*, http://www-misr.jpl.nasa.gov/.

NASA MODIS, 2014, *Moderate Resolution Imaging Spectrometer*, http://modis.gsfc.nasa.gov/.

NASA NPOESS, 2014, *National Polar Orbiting Operational Environmental Satellite System*, http://science.nasa.gov/missions/npoes/.

NASA NPP, 2011, *NPOESS Preparatory Project – Building a Bridge to a New Era of Earth Observations*, Washington: NASA, 20 p.

NASA/Orbimage SeaWiFS, 2014, *Sea-viewing Wide Field-of-view Sensor*, http://oceancolor.gsfc.nasa.gov/SeaWiFS/.

NASA Shuttle Photography, 2014, *NASA Space Shuttle Earth Observation Photography*, http://spaceflight.nasa.gov/gallery/images/shuttle/index.html.

NASA VIIRS, 2014, *Visible/Infrared Imager/Radiometer Suite*, http://npp.gsfc.nasa.gov/viirs.html.

NOAA AVHRR, 2014, *Advanced Very High Resolution Radiometer*, http://noaasis.noaa.gov/NOAASIS/ml/avhrr.html.

NOAA GOES, 2014, *Geostationary Operational Environmental Satellite*, http://www.oso.noaa.gov/goes/.

NOAA GVI, 2014, *NOAA Global Vegetation Index User's Guide*, www.ncdc.noaa.gov/oa/pod-guide/ncdc/docs/gviug/index.htm.

Pictometry ChangeFindr, 2014, *Pictometry ChangeFindr Reports*, NY: Pictometry International, Inc., http://www.pictometry.com.

Pictometry International, 2014, *ChangeFindr,* NY: Pictometry International, Inc., http://www.pictometry.com.

Puglia, S., Reed, J., and E. Rhodes, 2004, *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images: For the Following Record Types – Textual, Graphic Illustrations/Artwork/Originals, Maps, Plans, Oversized, Photographs, Aerial Photographs, and Objects/Artifacts*, Washington: National Archives and Records Administration, 87 p. http://www.archives.gov/preservation/technical/guidelines.pdf.

Raber, S., 2012, "Photo Science Teams with RapidEye forming an International Geospatial Partnership," Lexington, KY: Photo Science, www.sraber@photoscience.com.

RapidEye, 2014, *RapidEye High Resolution Satellite Imagery*, Brandenburg an der Havel, Germany: RapidEye, Inc., www.rapideye.net.

Solomonson, V., 1984, "Landsat 4 and 5 Status and Results from Thematic Mapper Data Analyses," *Proceedings*, *Machine Processing of Remotely Sensed Data*, W. Lafayette, IN: Lab for the Applications of Remote Sensing, 13–18.

SPOT Image, 2014, *SPOT 1 to SPOT 5*, Paris: Airbus Defense and Space, http://www.astrium-geo.com/en/4388-spot-1-to-spot-5-satellite-images.

SPOT Payload, 2014, *SPOT 6 and SPOT 7 Satellite Imagery*, Paris: Airbus Defense and Space, http://www.astrium-geo.com/en/147-spot-6-7-satellite-imagery.

USGS AVHRR, 2014, *AVHRR Spectral Channels, Resolution, and Primary Uses*, http://ivm.cr.usgs.gov/tables.php.

USGS Global Landcover, 2014, *Global Landcover*, http://landcover.usgs.gov/glcc/index.php.

USGS, 2010, *Landsat Data Continuity Mission Fact*, # Fact Sheet 2007-3093 Revised June, 2010, Reston: U.S. Geological Survey, 4 p.

USGS, 2013, *Acquiring Commercial Satellite Imagery from the National Geospatial-Intelligence Agency through the U.S. Geological Survey*, Washington: U.S. Geological Survey and Civil Applications Committee, February, 7 p.

Vane, G. and A. F. H. Goetz, 1993, "Terrestrial Imaging Spectrometry: Current Status, Future Trends," *Remote Sensing of Environment*, 44:117–126.

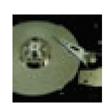
Wiechert, A., 2009a, "BING maps – A Virtual World," *Photogrammetric Week*, 2009, Berlin: Vexcel Imaging GmbH, 79 p.

Wiechert, A., 2009b, "Photogrammetry Versus LiDAR: Clearing the Air," *Photogrammetric Week*, 2009, July 22, Berlin: Vexcel Imaging GmbH, 79 p.

3 Digital Image Processing Hardware and Software



Analog (hard-copy) and digital remotely sensed data are used operationally in many Earth science, social science, and planning applications (e.g., Jensen and Jensen, 2013). Analog remotely sensed data such as positive 9 × 9 in. aerial photographs are routinely analyzed using the fundamental analog elements of image interpretation (e.g., size, shape, shadow) and optical instruments such as stereoscopes and zoom-transfer-scopes. Digital remote sensor data are analyzed using a digital image processing system that consists of computer hardware and special-purpose image processing software.

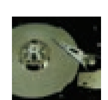


Overview

This chapter begins with a review of computer hardware characteristics that are of value when conducting digital image processing, including: type of computer, central processing unit (CPU), system random-access (RAM) and read-only (ROM) memory, mass storage and data archive considerations, video display spatial and spectral resolution, input and output devices, etc. The hardware associated with typical digital image processing laboratories is discussed.

High-quality digital image processing software is critical for successful digital image processing. The software should be easy to use and functional. The most important digital image processing functions are introduced. Many of the most commonly used digital image processing systems are then reviewed including their functional strengths. Image processing system cost constraints are introduced.

High-quality digital image processing software has significant capabilities. However, there is often a need to perform some type of geospatial analysis that is not available in the standard image processing software. When this situation occurs, a well-trained digital image processing professional may be able to program new geospatial code that can function within the standard image processing software. Digital image processing software that allows users to create custom digital image processing applications is identified.



Digital Image Processing Hardware Considerations

To successfully process digital remote sensor data it is usually best to hire people who are trained in a systematic body of knowledge (e.g., forestry, agronomy, urban planning, geography, geology, marine science) and who have considerable knowledge in GIScience (cartography/geodesy, remote sensing, and geographic information systems). They must understand the theoretical basis of remote sensing data-collection systems and the various digital image processing (and GIS) algorithms and how to properly apply the technology to their specific systematic body of knowledge. It is the quality and creativity of the people, not the hardware or the software, that dictates how useful the digital image processing will be.

Qualified image analysts select an appropriate digital image processing system that a) has a reasonable learning curve and is relatively easy to use, b) has a reputation for producing accurate results (ideally the software vendor has ISO certification), c) will produce

the desired results in an appropriate format (e.g., map products in a standard cartographic data structure compatible with most GIS), and d) is within their department's budget. Table 3-1 summarizes some of the important factors to consider when selecting a digital image processing system. It is useful to review these factors.

Central Processing Unit Considerations

Many Earth resource analysis and planning projects require that large geographic areas be inventoried and monitored through time. Therefore, it is common to obtain remotely sensed data of the study area. Unfortunately, the amount of digital data generated by a remote sensing system can be daunting. For example, a Landsat 5 Thematic Mapper 185 × 170 km scene consists of seven bands of 30 × 30 m data (the thermal channel was actually 120 × 120 but was resampled to 30 × 30 m) and is approximately 244 Mb (5,666 rows × 6,166 columns × 7 bands / 1,000,000 = 244 Mb). Change detection studies require multiple dates of remote sensor data (Canty, 2014).

Furthermore, many public agencies and Earth resource scientists are now taking advantage of even higher spatial and spectral resolution remote sensor data. For example, a single 11 × 11 km GeoEye, Inc. IKONOS-2 scene of panchromatic data consisting of 1 × 1 m pixels is 121 Mb (11,000 × 11,000 / 1,000,000 assuming 8-bit pixels). A single 512 × 512 pixel AVIRIS hyperspectral subscene contains 224 bands of 12-bit data and is 88.1 Mb (512 × 512 × 1.5 × 224 / 1,000,000). Processing such large remote sensor datasets requires a significant number of computations. The type of computer selected dictates how fast (efficient) the computations or operations can be performed and the precision with which they are made.

History of Central Processing Units and Efficiency Measurement

The central processing unit (CPU) is the computing part of the computer. It consists of a control unit and an arithmetic logic unit. The CPU:

- performs numerical integer and/or floating point calculations, and
- directs input and output from and to mass storage devices, color monitors, digitizers, plotters, etc.

The CPU's efficiency can be measured in terms of

- the number of cycles it can process in one second, e.g., 3.7 GHz means the CPU performs approximately 3.7 billion cycles per second,
- how many millions of instructions it can process per second (MIPS), e.g., 500 MIPS, and/or
- the number of transistors used by the CPU.

MIPS are difficult to compare between CPU architectures. In fact, MacNeil (2004) suggests that the term MIPS should probably be titled "misleading indicator of processor speed." Currently, the most reliable measure of the speed of a CPU is the number of transistors associated with the CPU.

To appreciate the quality of the computers that we routinely use for digital image processing today, it is instructive to briefly review the history of the central processing unit and consider some of its characteristics. The ENIAC was the first computer. It was invented in 1946 and weighed approximately 30 tons. In 1968, there were only 30,000 computers in the entire world—mostly mainframes that occupied entire rooms and refrigerator-sized mini-computers. People programmed the computers using punch cards. This changed when several people who worked for Fairchild Semiconductor left to start their own business—Intel, Inc.

TABLE 3–1 Factors to consider when selecting a digital image processing system.

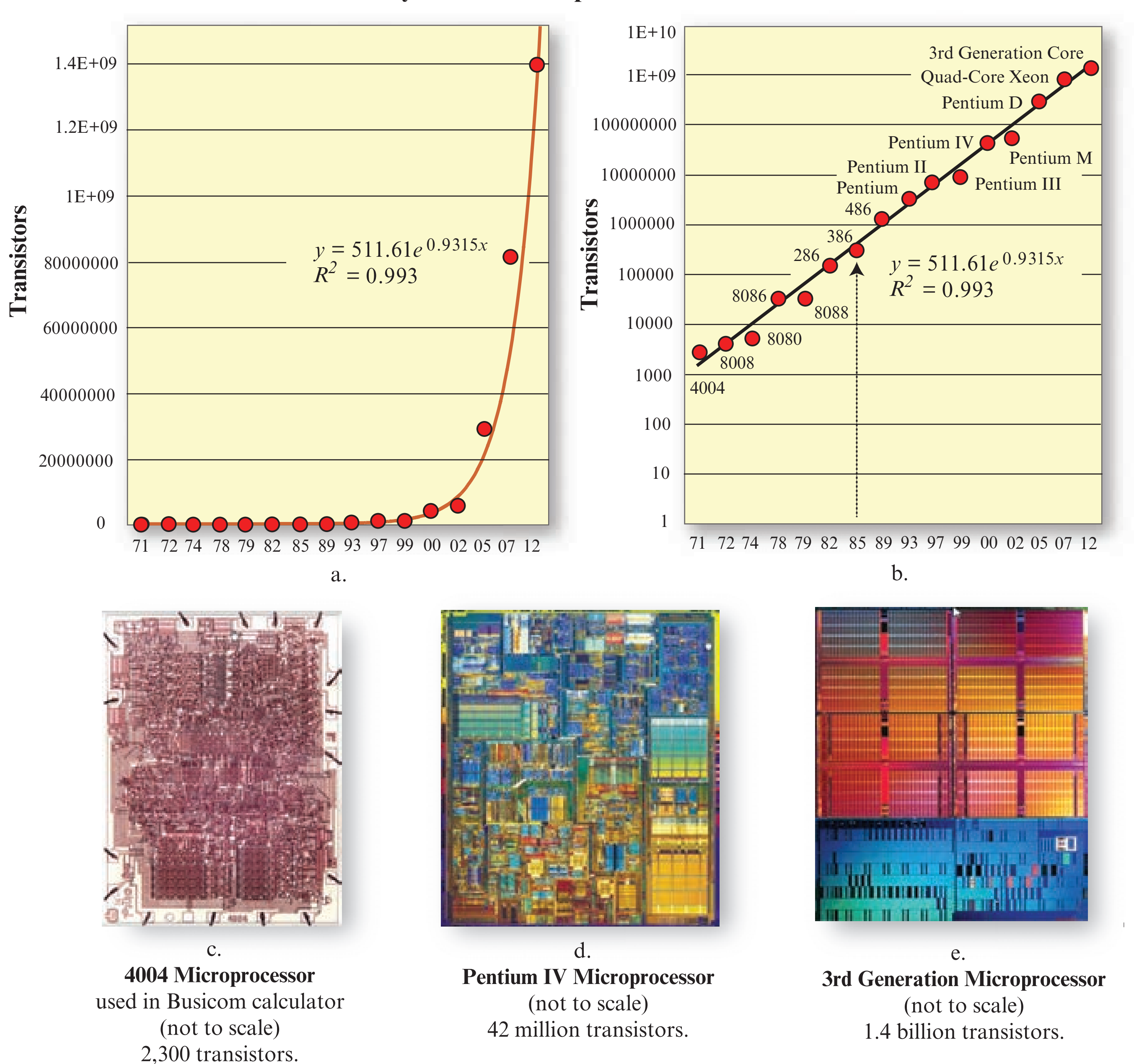
Digital Image Processing System Considerations
• Number and speed of the computer's CPU(s)
• Operating system (e.g., Microsoft Windows, UNIX, Linux, Apple)
• Amount of random-access memory (RAM), Video RAM, and graphics processors
• Number of image analysts that can use the system at one time and the mode of operation (e.g., interactive or batch)
• Serial and/or parallel processing capability
• Arithmetic coprocessor or array processor
• Software compiler(s)
• Type of mass storage (e.g., hard disk, CD, DVD, flash drive) and amount (e.g., gigabytes)
• Monitor display spatial resolution (e.g., 1024 × 768 pixels)
• Monitor color resolution (e.g., 24 bits of image processing video memory yields 16.7 million displayable colors)
• Input devices (e.g., optical-mechanical drum or flatbed scanners, area array digitizers)
• Output devices (e.g., CD, DVD, film-writers, ink-jet printers, dye sublimation printers)
• Networks (e.g., local area, wide area, Internet)
• Image processing software (e.g., ERDAS Imagine, Environment for Visualizing Images, PCI Geomatica, ER MAPPER, IDRISI, Esri Image Analyst, *e*Cognition, TNTmips, VIPER)
• Interoperability with major GIS software

TABLE 3–2 Historical development of Intel, Inc., central processing units for personal computers (Intel, 2014).

Central Processing Unit (Date introduced)	Clock Speed (KHz) (MHz) (GHz)	Transistors	Manufacturing Technology (μm) (nm)	Significance
4004 (1971)	108 KHz	2,300	10 μm	First microcomputer chip
8008 (1972)	500–800 KHz	3,500	10 μm	2× as powerful as the 4004
8080 (1974)	2 MHz	4,500	6 μm	Made video games and home computers possible
8086 (1978)	5 MHz	29,000	3 μm	First 16-bit processor
8088 (1979)	5 MHz	29,000	3 μm	Used in first IBM PC
286 (1982)	6 MHz	134,000	1.5 μm	3–6× performance of 8086
386 (1985)	16 MHz	275,000	1.5 μm	First to process 32-bit data and run multiple programs at one time
486 (1989)	25 MHz	1.2 million	1 μm	First integrated floating point unit
Pentium (1993)	66 MHz	3.1 million	0.8 μm	Could now do speech, sound
Pentium II (1997)	300 MHz	7.5 million	0.25 μm	MMX media enhancement
Pentium III (1999)	500 MHz	9.5 million	0.18 μm	Low-power states conserve energy
Pentium IV (2000)	1.5 GHz	42 million	0.18 μm	Ushered in nanotechnology age
Pentium M (2002)	1.7 GHz	55 million	90 nm	Revolutionized portable computing
Pentium D (2005)	3.2 GHz	291 million	65 nm	First desktop duel-core processors
Quad-Core Xeon (2007)	3.0 GHz	820 million	45 nm	Micro architecture enhancements
3rd and 4rd Generation Core (2013–2014)	3–4 GHz	>1.4 billion	22 nm	Hyper-threading technology

In 1969, Nippon Calculating Machine Corporation asked Intel, Inc., to design computer chips for its new Busicom printing calculator. Intel designed a set of four chips known as the MCS-4. It included a central processing unit (CPU) chip—the 4004—as well as a supporting read-only memory (ROM) chip for the custom applications programs, a random-access memory (RAM) chip for processing data, and a shift-register chip for the input/output (I/O) port. Intel subsequently purchased the rights from Nippon Calculating Machine Corporation and launched the Intel® 4004 processor and its chipset with an advertisement in the November 15, 1971, issue of *Electronic News* (Table 3-2). The Intel® 4004 became the first general-purpose programmable CPU in the marketplace—a "building block" that engineers could purchase and then customize with software to perform different functions in a wide variety of electronic devices (Intel, 2014).

Gordon Moore was one of the founders of Intel, Inc. As early as April 19, 1965, he published an article in *Electronics Magazine* in which he made a prediction about the semiconductor industry. He realized that each new chip contained roughly twice as much capacity as its predecessor and each chip was released within 18 to 24 months of the previous chip. If this trend continued, he reasoned, computing power would rise exponentially over relatively brief periods of time.



c.
4004 Microprocessor
used in Busicom calculator
(not to scale)
2,300 transistors.

d.
Pentium IV Microprocessor
(not to scale)
42 million transistors.

e.
3rd Generation Microprocessor
(not to scale)
1.4 billion transistors.

FIGURE 3–1 Moore's law suggests a relationship between the capacity of computer chips (e.g., number of transistors) and the 18- to 24-month release interval. a,b) Number of transistors in Intel microprocessors displayed in real numbers and exponentially. c) The 108-KHz 4004 Intel chip produced in 1971 had 2,300 transistors. d) The 1.5-GHz Pentium IV produced in 2000 had 42 million transistors. e) The 3.7-GHz 3rd Generation processors have more than 1.4 billion processors (information courtesy of Intel, 2014). It is important to note that Intel, Inc., is not the only manufacturer of computer chips.

Moore's law described a trend that has continued and is still remarkably accurate. For example, the number of transistors in Intel chips from 1971 through 2012 listed in Table 3-2 ranged from 2,300 to more than 1.4 billion (Figure 3-1a). The same information is plotted logarithmically in Figure 3-1b revealing a straightforward linear pattern. Photographs of the 4004, Pentium IV, and 3rd Generation CPUs are shown in Figures 3-1c–e.

Many personal computers, workstations, and especially mainframe computers have multiple CPUs. A single CPU functions in a serial manner. The presence of multiple CPUs allows parallel processing to take place. The operating system and/or GIS software is able to parse (distribute) different tasks to the several CPUs, dramatically improving processing speed.

Personal Computers are Ideal for Digital Image Processing and GIS Analysis



FIGURE 3–2 Properly configured personal computers and computer workstations are ideal for digital image processing and GIS analysis. In this example, the outline (and area) of several large buildings (shown in red) have been extracted from the natural color orthophoto (data courtesy State of Utah).

Type of Computer

All digital image processing system software vendors list the minimum system requirements necessary to run their software. Keep in mind that these are the minimum requirements. Computers with components that exceed the minimum requirements generally will be more useful and efficient when conducting digital image processing. If the computer has only the minimum hardware required to run the digital image processing software, the software will probably run slowly and will most likely dominate the operating system's resources. This is especially a problem if you are a "multi-tasker" who frequently has other programs running at the same time as the digital image processing software (e.g., word processor, graphics program, Internet browser).

Computers used for digital image processing can be organized into three main categories: personal computers (PCs), computer workstations, and mainframe computers.

Personal Computers

Personal computers (PCs) are the workhorses of the digital image processing industry (Figure 3-2). These include relatively inexpensive computers such as desktops, laptops, and tablets. Most of the personal computers now come with multiple CPUs. A **multi-core processor** is a single computing component with two or more independent central processing units (called **cores**), that read and execute program instructions. CPUs were originally developed with only one core. A dual-core processor has two cores (e.g., Intel Core Duo), a quad-core processor contains four cores (e.g., Intel quad-core *i*7), a hexa-core processor contains six cores (e.g., Intel Core *i*7 Extreme Edition), an octa-core processor contains eight cores and so on.

Current PCs have much greater clock speeds, billions of transistors and can process instructions faster than their predecessors because they have 64-bit registers (word size) compared to the 8-bit registers used historically. Many digital image processing–related companies provide their employees with high-quality personal computers because of their low initial cost and inexpensive maintenance requirements. Interestingly, it seems that one can always purchase a "good" personal computer for digital image processing analysis for under $2,500. Ideally, the computer should have >8 GB of RAM, a large hard disk (>1 TB), a rewritable disk drive (e.g., DVD-RW), a precise graphic input device (e.g., a cursor), and a good graphics display system (monitor and video card). Common operating systems suitable for personal computers include Microsoft Windows products (e.g., Windows 8), UNIX, Linux, and the Apple Macintosh OS. These operating systems allow computers to be networked and have access to the Internet.

Computer Workstations

Computer workstations usually contain more powerful processors, more RAM, larger hard disk drives, and very high-quality graphics display capability. These improved components allow workstations to perform digital image processing analysis more rapidly than a personal computer. However, the cost of a workstation is usually two to three times more than the cost of a personal computer. The most common workstation operating systems are UNIX, Linux, and various Microsoft Windows products.

Mainframe Computers

Mainframe computers perform calculations more rapidly than personal computers and workstations, and they are able to support hundreds of users simultaneously. They may contain hundreds of CPUs in which case they are usually referred to as **super computers**. Mainframes are ideal for intensive CPU-dependent tasks such as overlay analysis, large database operations, and raster rendering. If desired, the output from mainframe processing can be passed to a personal computer or workstation for less intensive processing. Mainframe computers are usually expensive to purchase and maintain. Also, digital image processing software for mainframe computers is more expensive.

Read-Only Memory and Random-Access Memory

Computers used for digital image processing have several other characteristics that must be considered, such as **Read-Only Memory** (**ROM**) and **Random-Access Memory** (**RAM**). ROM retains information even after the computer is turned off because power is supplied by a battery that occasionally must be replaced. When a computer is turned on, the computer examines the information stored in the various ROM registers and uses this information to proceed. Most personal computers have sufficient ROM to perform quality digital image processing.

RAM is the computer's primary *temporary* workspace. Unlike ROM, the data stored in RAM are lost when the computer is turned off. Computers should have sufficient RAM for the operating system, image processing software, and any spatial data that must be held in temporary memory while calculations are performed. Because of this, the amount of RAM is one of the most important considerations when purchasing a computer for digital image processing. It seems that one cannot have too much RAM for digital image processing. It is good practice to have > 8 GB of RAM for digital image processing applications. Fortunately, RAM prices continue to decrease as the amount of RAM and its speed increase.

Serial and Parallel Image Processing

Most PCs, workstations, and mainframe computers now have multiple CPUs that operate concurrently (Figure 3-3). Some expensive parallel super computers have hundreds and even thousands of cores (e.g., IBM Blue Gene/Q). Specially written parallel processing software can parse (distribute) the remote sensor data to specific CPUs (cores) to perform digital image processing (e.g., Sorokine, 2007). This can be much more efficient than processing the data serially. For example, consider performing a per-pixel classification on a typical 1,024 row by 1,024 column remote sensing dataset (Figure 3-3a). In the first example, each pixel is classified by passing the spectral data associated with each pixel to the CPU, classifying it, and then progressing to the next pixel. This is inefficient *serial* processing (Figure 3-3a).

Conversely, suppose that instead of just one CPU we had 1,024 CPUs. In this case, the class of each of the 1,024 pixels in row one could be determined using 1,024 separate CPUs (Figure 3-3b). If this were the only task required in the procedure, the *parallel* image processing would classify the line of data approximately 1,024 times faster than serial processing. In an entirely different parallel configuration, each of the 1,024 CPUs could be allocated an entire row of the dataset.

While parallel processing might appear to be the panacea to digital image processing efficiency, it is important to note that the potential speedup of an algorithm using parallel computing is constrained by **Amdahl's law** (Amdahl, 1967). It states that if α is the fraction of running time a program spends on *non*-parallel parts (called sequential parts), then the maximum speedup of the process (S) using parallel code is:

$$S = \frac{1}{\alpha}. \tag{3.1}$$

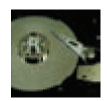
For example, if the non-parallel (sequential) portion of a program accounts for 10% of the running time, we can get no more than a 10× speedup, regardless of how many processors are added. This is demonstrated in Figure 3-3c:

$$S = \frac{1}{0.10} = 10\times$$

If the non-parallel (sequential) portion of a program accounts for 50% of the running time, we can get no more than a 2× speedup, regardless of how many processors are used as shown in Figure 3-3c:

$$S = \frac{1}{0.50} = 2\times$$

Interestingly, all of the CPUs (cores) do not have to reside on the same computer or even in the same city to perform parallel processing. It is possible to perform parallel processing by connecting individual computer systems via a network. This type of parallel processing requires sophisticated distributed processing software. In practice, it is difficult to parse a program so that multiple CPUs can efficiently execute different portions of the program without interfering with one another. Vendors continue to develop digital image processing code that takes advantage of multiple-core parallel architecture.



Mode of Operation and User Interface

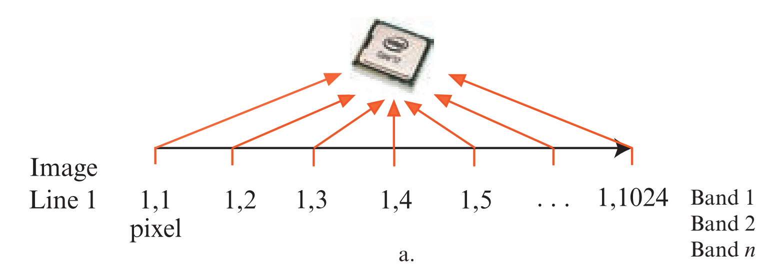
Image analysts can process remotely sensed data interactively or in batch mode. Ideally the processing takes place in an interactive environment using a well-crafted graphical user interface (GUI).

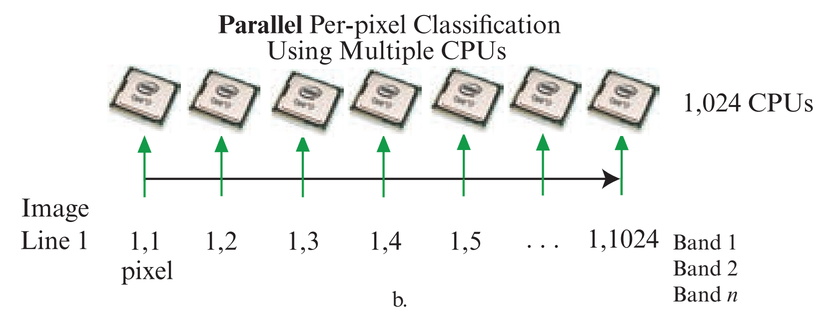
Mode of Operation

Breakthroughs in analyst image understanding and scientific visualization are generally accomplished by

Serial versus Parallel Digital Image Processing to Perform Per-pixel Classification

Serial Per-pixel Classification
Using a Single CPU





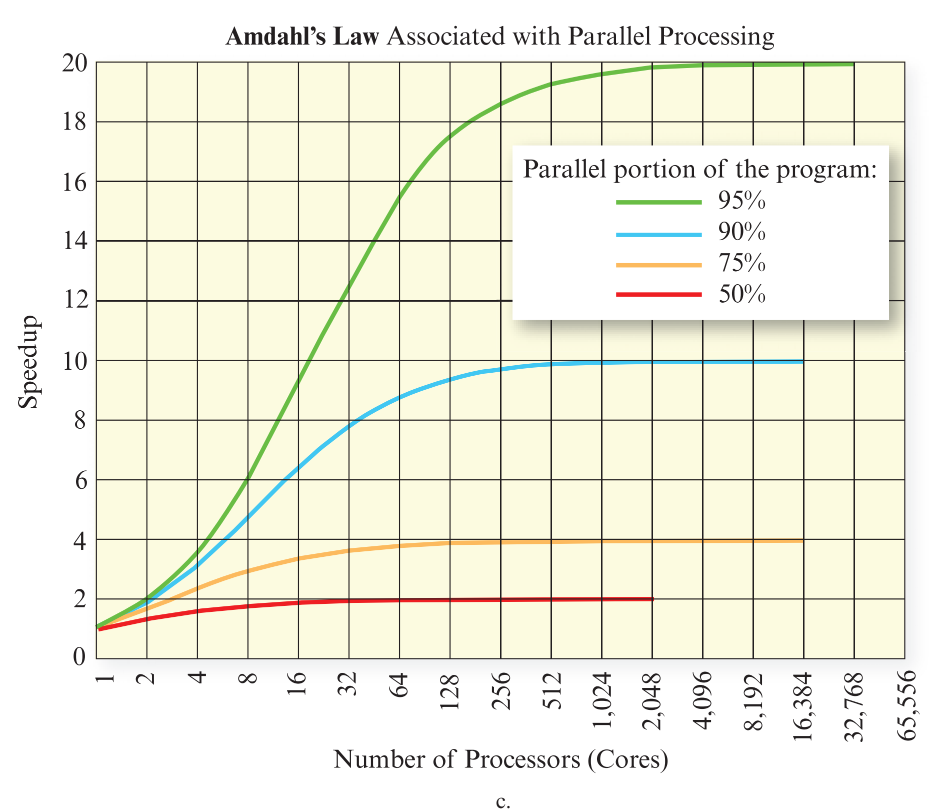


FIGURE 3–3 Digital image classification using a) serial, and b) parallel image processing logic. There are 1,024 pixels in the line of data. In a serial processing environment, each pixel of each line in each band is processed sequentially using a single CPU. In a parallel processing environment, each of the 1,024 CPUs could be requested to process 1) each individual pixel in a line of remote sensor data containing 1,024 pixels (as shown), or 2) an entire line of data (not shown). c) A graphic representation of Amdahl's Law documenting the amount of speedup expected when a computer program consists of certain proportion of parallel code. For example, a program that consists of 10% sequential code and 90% parallel code can have a maximum speedup of 10×.

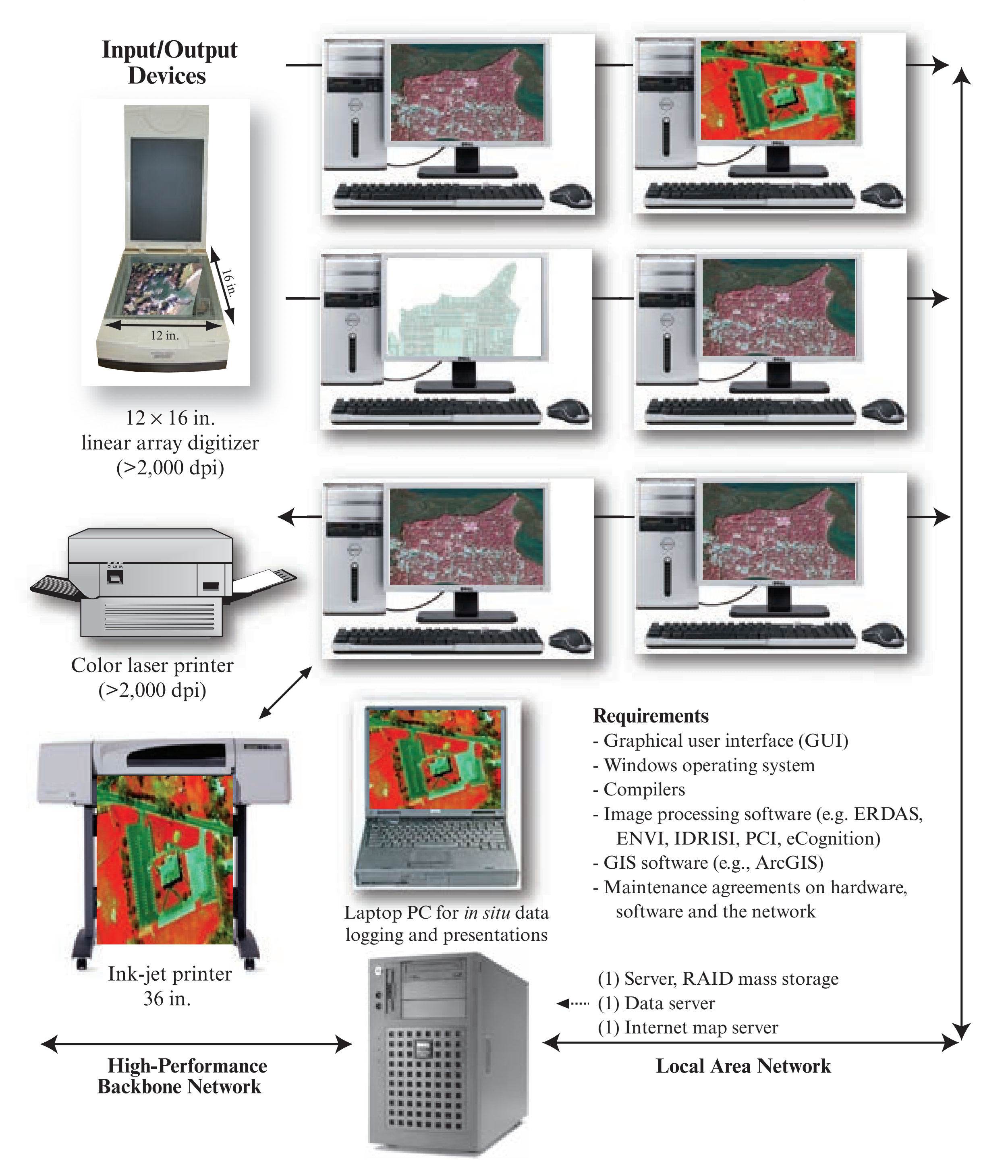


FIGURE 3–4 Digital image processing labs usually contain a number of relatively sophisticated PCs. In this example, there are six 24-bit color desktop personal computers (PCs), one laptop, and a server. The PCs communicate locally via a local area network (LAN) and with the world via the Internet. Each 4 GHz PC has >8 GB random-access memory, >1 TB hard disk space, and a CD, DVD and/or blue-ray disk. Image processing software and remote sensor data can reside on each PC (increasing the speed of execution) or be served by the server, minimizing the amount of mass storage required on each PC. Scanners, printers, and plotters are required for input of digital remote sensor data and output of important results.

Typical Digital Image Processing and GIS Computer Laboratory



FIGURE 3–5 This typical digital image processing and GIS computer laboratory consists of multiple high-end personal computers with single, very fast CPUs, >1 TB mass storage, high resolution computer screens (1,900 × 1,200) with 512 MB to 1 GB of video memory, cursors, connection to a local area network (LAN) and access to the Internet. All of the computers have access to the color laser printer and scanner and to several devices not shown including a digitizing table and E-size ink-jet plotter. The digital overhead projector allows the instructor to display items of interest and perform real-time GIS analysis.

placing the analyst as intimately in the image processing loop as possible and allowing his or her intuitive capabilities to take over. Ideally, every analyst has access to his or her own digital image processing system. Unfortunately, this is not always possible due to cost constraints. The sophisticated PC or workstation laboratory environment shown in Figure 3-4 would be ideal for six or seven people doing digital image processing in an educational or research environment. It would probably be ineffective for education or short course instruction where many analysts (e.g., >20) must be served.

A real digital image processing/GIS computer laboratory is shown in Figure 3-5. In this laboratory, the computers are connected via a local area network (LAN) that allows multiple users to access spatial data from a common source (e.g., a file server). All of the laboratory computers are connected to the Internet which allows users to efficiently download geospatial data from inside or outside the university and communicate with one another. All the computers have access to high-quality input (e.g., coordinate digitizing table, scanner) and output devices (e.g., E-size plotter, color laser printer). Each computer has access to >5 TB of mass storage and the ability to back up important files locally or at a common repository. Whiteboards are present to list tasks and promote discussion among the analysts. The digital overhead projector allows the instructor to demonstrate concepts and interactively perform image processing. It is also possible to perform digital image processing over the Internet and access data in a "cloud."

Interactive Graphical User Interface

One of the best scientific visualization environments for the analysis of remote sensor data takes place when the analyst communicates with the digital image processing system interactively using a point-and-click **graphical user interface (GUI)**. Most sophisticated image processing systems are configured with a friendly, point-and-click GUI that allows rapid display of images and the selection of important image processing functions. Easy to use digital image processing graphical user interfaces are found in:

- Exelis Inc., Environment for Visualizing Images (ENVI®) (Figure 3-6),

ENVI Graphical User Interface (GUI)

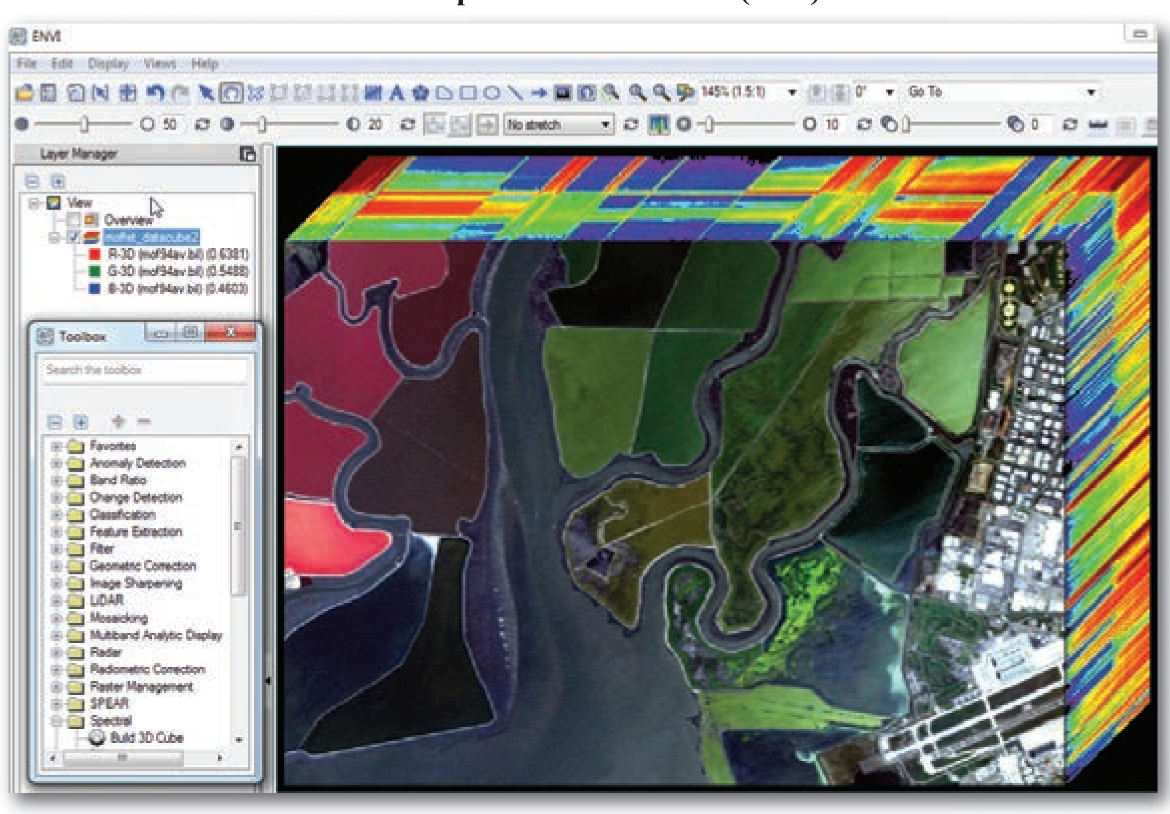


FIGURE 3–6 The Exelis, Inc., Environment for Visualizing Images (ENVI®) graphical user interface. The "Build 3D Cube" program is being used to create a 3D datacube of AVIRIS hyperspectral imagery of wetland and salt evaporation ponds near Moffett Field, CA. There are 300 rows, 500 columns, and 56 of 224 AVIRIS bands in this particular dataset. Bands 25 (center at 0.6381 μm; red), 16 (0.5488 μm; green), and 7 (0.4603 μm; blue) are displayed on top of the cube in a natural color composite (interface courtesy of Exelis, Inc.; AVIRIS data courtesy of NASA Jet Propulsion Laboratory).

- Intergraph, Inc., ERDAS IMAGINE® and Leica Photogrammetry Suite (LPS®) (Figure 3-7),
- Clark University's IDRISI®,
- Purdue University's MultiSpec®, and others.

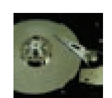
Several geographic information systems perform digital image processing and use effective GUIs, including:

- Esri, Inc., ArcGIS® image Classification, Feature Analyst® and Spatial Analyst® interfaces (Figure 3-8), and
- GRASS GIS®.

Adobe, Inc., Photoshop® has a time-tested GUI used by thousands of people each day. The software is especially useful for enhancing photographs and images that have three or fewer bands of data and geometric rectification and classification are not required.

Batch Processing

Non-interactive, batch processing can be of value for time-consuming processes such as image rectification, segmentation, mosaicking, orthophoto generation, and special filtering. Batch processing frees up laboratory PCs or workstations during peak demand because the time-consuming jobs can be put in a que and executed when the computer is otherwise idle (e.g., during early morning hours). Batch processing can also be useful during peak hours because it allows the analyst to set up a series of operations that can be executed in sequence without operator intervention.



Computer Operating System and Compiler(s)

The **operating system** is the first program loaded into memory (RAM) when the computer is turned on. The operating system controls all of the computer's higher order functions and resides in RAM at all times. The operating system provides the user interface, controls multitasking, handles input and output to the hard disk and all peripheral devices such as DVDs, scanners,

ERDAS Imagine Graphical User Interface (GUI)

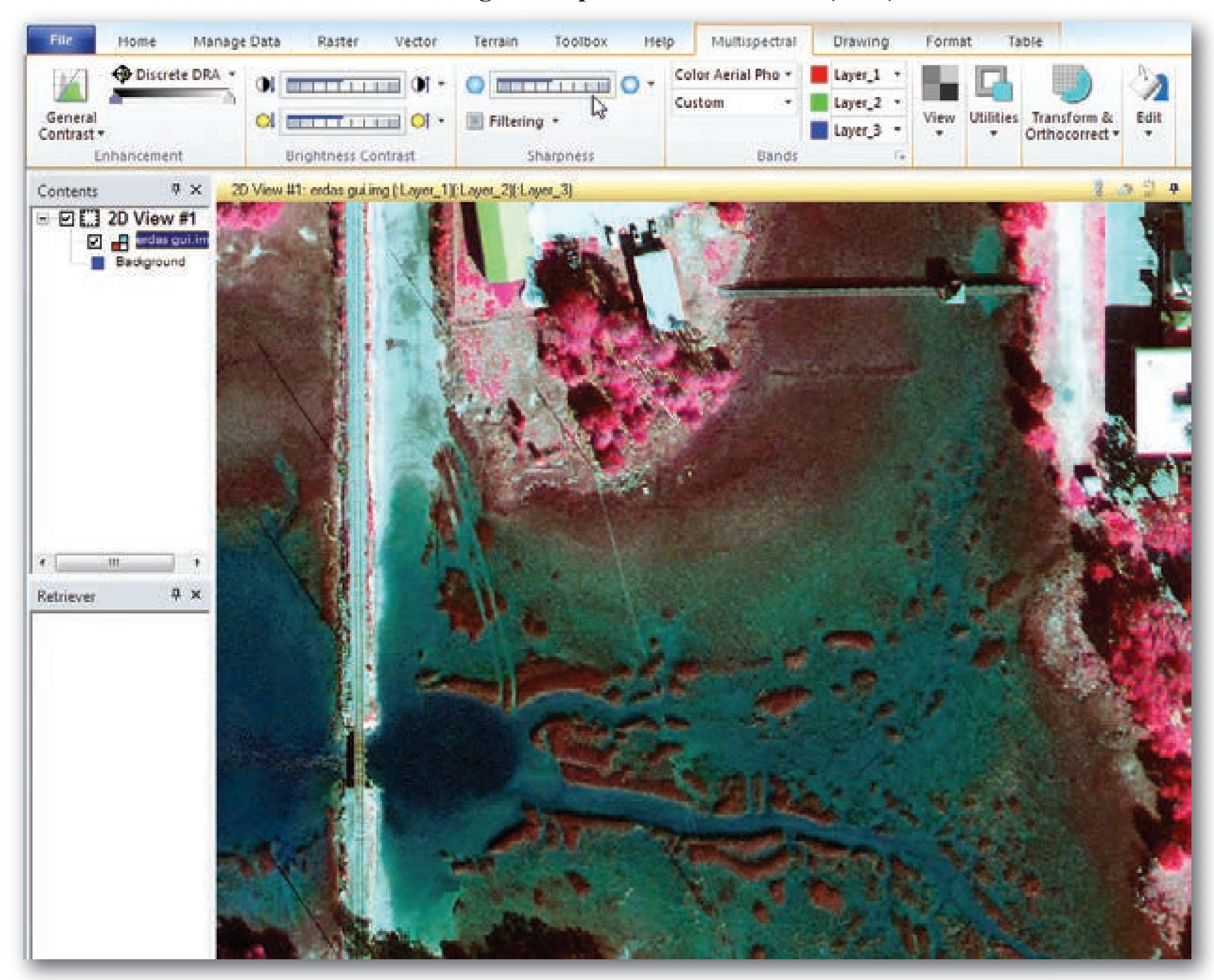


FIGURE 3–7 The Intergraph ERDAS Imagine® graphical user interface with the "Multispectral" pallet highlighted. The "sharpness" of this image near Beaufort, SC, is in the process of being adjusted (user interface courtesy of Intergraph, Inc., part of Hexagon, Inc.; image courtesy of Dan Morgan, Beaufort County GIS Department).

printers, plotters, and the color display. All digital image processing software must communicate with the operating system.

Different types of digital image processing software are able to run on most major operating systems (e.g., Windows, UNIX, Linux, Mac OS; Tables 3-1 and 3-5). Users should select digital image processing software that runs on an operating system that they are comfortable with. This helps reduce the learning curve as you begin working with the image processing software. If you are going to run the applications software on your personal computer, be aware that you may have to upgrade to a new or different operating system in order to run the image processing software. If you are installing the image processing software on an existing computer you must also be sure that the computer has sufficiently fast CPU(s), enough memory (RAM), and adequate hard disk space for the image processing software to run efficiently.

Input Devices

As discussed in Chapter 2, it is important to have a quality scanner to digitize hard-copy imagery such as panchromatic, natural color, and/or color-infrared aerial photography. Aerial photographs are typically 9 × 9 in. in size. Therefore, it is important to have a scanner with an effective area of at least 12 × 16 in. so that an entire 9 × 9 in. aerial photograph can be scanned in one pass (Figure 3-4).

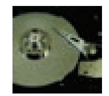
Output Devices

A GIS should be able to output high-quality maps, images, charts, and diagrams in both small (e.g., A-size

ArcGIS Classification Graphical User Interface (GUI)

FIGURE 3–8 The Esri ArcGIS ArcMap® graphical user interface with the image Classification, Feature Analyst®, and Spatial Analyst® extensions turned on. The classification user interface is being used to collect training samples of water and wetland adjacent to Beaufort, SC (user interface courtesy of Esri, Inc.; image courtesy of Dan Morgan, Beaufort County GIS Department).

plots 8.5 × 11 in.) or large formats (e.g., E-size plots 36 × 48 in.). To accomplish this, small- and large-format printers are required. Inexpensive ink-jet or color laser printers can be used for small format printing, and E-sized ink-jet plotters can be used for large formats.



Data Storage and Archiving Considerations

Digital image processing of remote sensing and related GIS data requires substantial mass storage resources. Therefore, the mass storage media should allow relatively rapid access times, have longevity (i.e., last for a long time), and be inexpensive.

Rapid Access Mass Storage

The best way to make digital remote sensor data rapidly available to CPU(s) is to place the data on a hard disk, compact disk (CD), digital video disk (DVD), flash drive, and/or in the Internet "cloud" where each pixel of the data matrix may be accessed at random (not serially) and at great speed (i.e., within microseconds). Because of the reduced cost of mass storage, it is common for digital image processing laboratories to have gigabytes of hard-disk mass storage associated with each PC or workstation. For example, each PC in the laboratory shown in Figure 3-5 has access to > 1 TB of hard disk mass storage. Some image processing laboratories use RAID (redundant arrays of inexpensive hard disks) technology in which two or more drives working together provide increased performance

and various levels of error recovery and fault tolerance. It is now common for scientists and students to use relatively inexpensive CDs, DVDs, and especially flash drives to back up their remote sensing data and results after every session. The cost of hard disk, CD, DVD, flash drive, and cloud data storage and access per gigabyte continues to decline dramatically, which is wonderful for remote sensing science and education.

Cloud computing is becoming a technology that almost every industry that provides or uses hardware, software, and data storage will utilize (Kouyoumjia, 2010; Yang and Huang, 2013). Cloud computing may eventually lead to a future where we do not compute on local computers. Rather, computing will occur on centralized facilities operated by third-party computer and data storage utilities. In a cloud computing environment, the data or software stored on remote servers is provided via the Internet to all types of computers.

Presently, there are two main types of cloud computing: public and private. In the public cloud, the infrastructure and services are owned and sold by a separate organization. Data space on the public cloud is available to all computer users. Private cloud computing is available for companies or organizations that are not comfortable placing their data files in a public cloud. A private cloud provides data storage in a remote location that is maintained behind a restrictive firewall.

Archiving Considerations—Longevity

Significant sums of money are spent purchasing remote sensor data by commercial companies, natural resource agencies, and universities. Storing massive amounts of original remote sensor data, processed remote sensor data, and project results is no trivial matter. Unfortunately, most of the time not enough attention is given to how the expensive data are archived to protect the long-term investment. Figure 3-9 depicts several types of analog and digital remote sensor data mass storage devices and the average time to physical obsolescence, that is, when the media begin to deteriorate and/or information is lost. Interestingly, properly exposed, washed, and fixed analog black-and-white aerial photography negatives have considerable longevity, often more than 100 years. That is why we still have pictures of Abraham Lincoln from the 1860s. Color negatives with their respective dye layers have longevity, but not as much as the black-and-white negatives. Black-and-white paper prints have greater longevity than color prints because the dyes in the color prints often begin to fade after a few years.

For all practical purposes, floppy disk and magnetic tape media are obsolete (Figure 3-9). The magnetic tape medium is particularly problematic because it stretches over time. The dimensional instability can make it very difficult to read the data stored on tape. Interestingly, there are thousands of remote sensing images acquired by SPOT, Landsat, and other sensors during the 1970s through 1990s stored on tape in laboratories throughout the world that need to be transferred to more robust storage media (e.g., DVD) before it is too late.

Hard disks, CDs, DVDs, and even flash drives provide long-term storage potential (>100 years) (Figure 3-9). However, it is very important to remember when archiving digital remote sensor data that sometimes it is the loss of the read–write software and/or the read–write hardware (the drive mechanism and heads) that becomes a problem and not the digital media itself (Rothenberg, 1995). Therefore, as new computers are purchased it is a good idea to archive all the computer system hardware and software that is representative of a certain computer era so that one can always read any data stored on old digital mass storage media. This includes the computer, hard-disk, monitor, keyboard, mouse, DVD, etc.

While cloud storage will continue to grow and evolve, there are some limitations. The data transfer rate between an individual computer and the cloud can be slow. This is a problem when large geospatial datasets must be accessed via the cloud. Privacy and security are two other concerns with cloud-computing. In some cases, customer or government data security issues may preclude storing sensitive data in the cloud (Kouyoumjia, 2010). Finally, cloud data storage can be a viable long-term data storage technology if the commercial firms managing the cloud remain in business and rigorously maintain the data through time.



Computer Display Spatial and Color Resolution

The display of remote sensor data on a computer screen is one of the most fundamental tasks of digital image analysis. Careful selection of the computer display characteristics will provide the optimum visual image analysis environment for the human interpreter. The two most important characteristics are computer display spatial and color resolution.

Computer Screen Display Resolution

The digital image processing system should be able to display at least 1,024 rows by 1,024 columns on the computer screen at one time. This allows larger geographic areas to be examined and places the terrain of interest in its regional context. Most Earth scientists

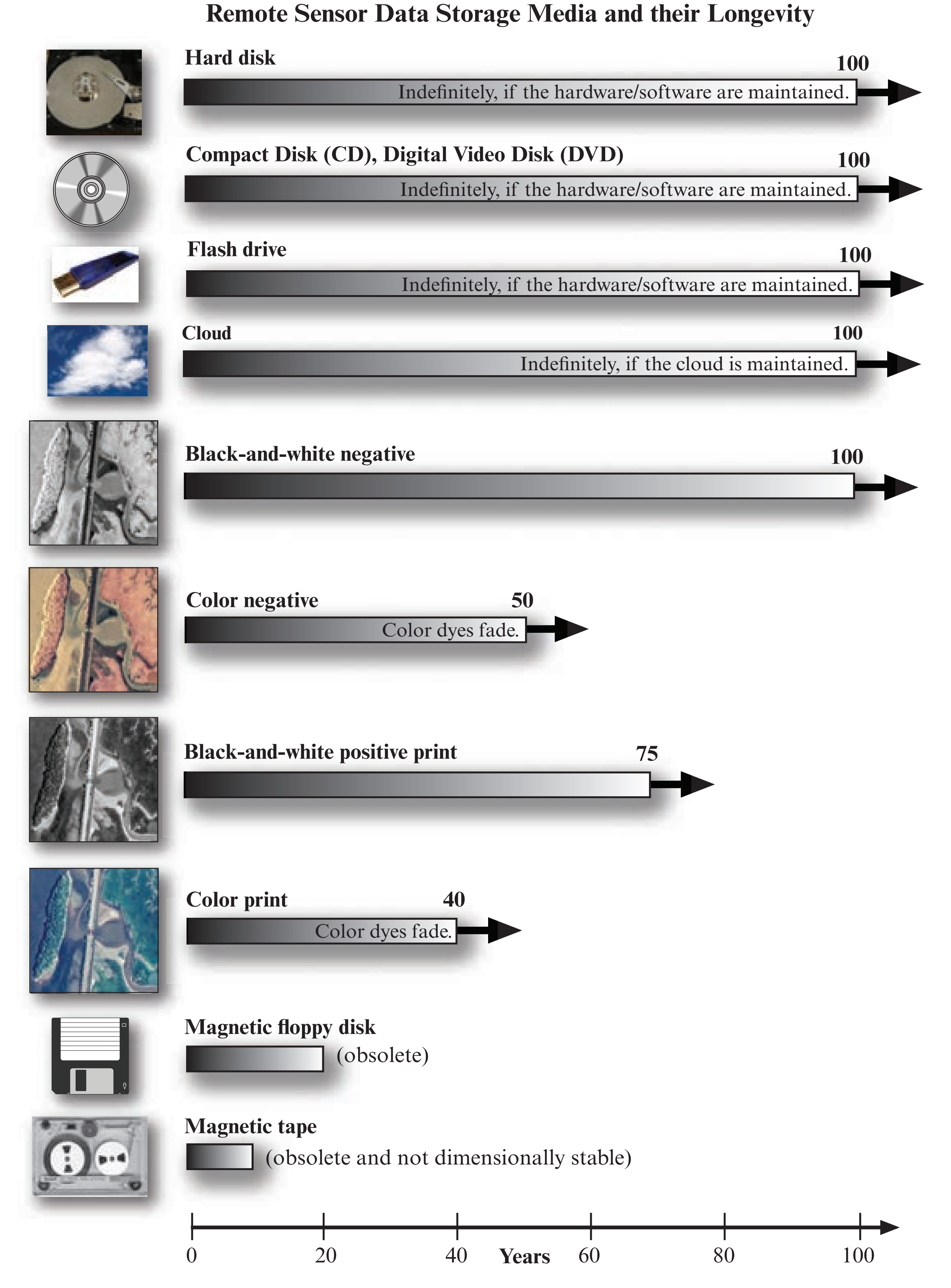


FIGURE 3–9 Different types of analog and digital remote sensor data mass storage alternatives and the average time to physical obsolescence. The loss of software or hardware to read the archived digital media is often the most serious problem.

prefer this regional perspective when performing terrain analysis using remote sensor data. It is disconcerting to have to analyze four 512 × 512 images when a single 1,024 × 1,024 display provides the information at a glance. An ideal screen display resolution is 1,600 × 1,200 pixels.

Computer Screen Color Resolution

The computer screen color resolution is the number of gray-scale tones or colors (e.g., 256) that can be displayed on the computer screen at one time out of a palette of available colors (e.g., 16.7 million). For many applications, such as high-contrast black-and-white linework cartography, only 1 bit of color is required (i.e., either the line is black or white [0 or l]). For more sophisticated computer graphics for which many shades of gray or color combinations are needed, up to 8 bits (or 256 colors) may be required. Most thematic mapping and GIS applications may be performed well by systems that display just 64 user-selectable colors out of a palette of 256 colors.

Conversely, the analysis and display of remote sensor image data require much higher screen color resolution than cartographic and GIS applications. For example, most relatively sophisticated digital image processing systems can display a tremendous number of unique colors from a large color palette (e.g., 16.7 million). The primary reason for these color requirements is that image analysts must often display a composite of several images at one time on the screen. This process is called **color compositing**. For example, to display a typical color-infrared image of Landsat Thematic Mapper data, it is necessary to composite three separate 8-bit images (e.g., the green band [TM 2 = 0.52 to 0.60 μm], the red band [TM 3 = 0.63 to 0.69 μm], and the reflective infrared band [TM 4 = 0.76 to 0.90 μm]). To obtain a true-color composite image that provides every possible color combination for the three 8-bit images requires that 2^{24} colors (16,777,216) be available in the palette (Table 3-3). Such true-color, direct-definition systems are relatively expensive because every pixel location must be bit-mapped. This means that there must be a specific location in memory that keeps track of the exact blue, green, and red color value for every pixel. This requires substantial amounts of computer memory which are usually present in what is called an image processor (discussed in the next chapter). Given the availability of image processor memory, the question is: what is adequate color resolution?

Generally, 4,096 carefully selected colors out of a very large palette (e.g., 16.7 million) seem to be the minimum acceptable for the creation of remote sensing color composites. This provides 12 bits of color, with 4 bits available for each of the blue, green, and red image planes (Table 3-3). For image processing applications other than compositing (e.g., black-and-white image display, color density slicing, pattern recognition classification), the 4,096 available colors and large color palette are adequate. However, the larger the palette and the greater the number of displayable colors at one time, the better the representation of the remote sensor data on the screen for visual analysis. More information about how images are displayed using an image processor is in Chapter 5. The network configured in Figure 3-4 has six 24-bit color workstations.

TABLE 3–3 Image processing memory required to produce various numbers of displayable colors on the computer screen.

Image Processor Memory (bits)	Maximum Number of Colors Displayable at One Time on the Screen
1	2 (black and white)
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	1,024
11	2,048
12	4,096
13	8,192
14	16,384
15	32,768
16	65,536
17	131,072
18	262,144
24	16,777,216

Several remote sensing systems now collect data with 10-, 11-, and even 12-bit radiometric resolution with brightness values ranging from 0 to 1023, 0 to 2047, and 0 to 4095, respectively. Unfortunately, despite advances in video technology, at the present time it is necessary to generalize (i.e., dumb down) the radiometric precision of the remote sensor data to 8 bits per pixel for display on the computer screen simply because current video display technology cannot handle the demands of the increased precision.



Digital Image Processing Software Considerations

The digital image processing software selected is critical to the successful completion of the image processing project or research. A good rule of thumb is to always use digital image processing software that has an excellent reputation. You may be making decisions based on the output from the image processing that can impact people, flora, and fauna. Your reputation will be dependent upon how carefully you structure the research question or application and the proper use of

algorithms and procedures within the image processing software.

Image Processing Functions

Many of the most important functions performed using digital image processing systems are summarized in Table 3-4. Personal computers now have the computing power to perform each of these functions. This textbook examines many of these digital image processing functions in subsequent chapters, including radiometric and geometric preprocessing (Chapters 6 and 7), enhancement (Chapter 8), information extraction (Chapters 9 through 11), change detection (Chapter 12), and accuracy assessment (Chapter 13).

It is not good for remotely sensed data to be analyzed in a vacuum. Remote sensing information fulfills its promise best when used in conjunction with other ancillary data (e.g., soils, elevation, and slope) often stored in a geographic information system (GIS). Therefore, the ideal digital image processing system should be able to process the digital remote sensor data as well as perform any necessary GIS processing (du Plessis, 2012). It is usually not efficient to exit the digital image processing system, log into a GIS system, perform a required GIS function, and then take the output of the procedure back into the digital image processing system for further analysis. Most integrated systems perform both digital image processing and GIS functions and consider map data as image data (or vice versa) and operate on them accordingly.

Most digital image processing systems have some limitations. For example, most systems can perform multispectral classification on a few bands of imagery, but only a few systems can perform hyperspectral analysis on hundreds of bands of imagery. Similarly, only a few systems can perform soft-copy photogrammetric operations on overlapping stereoscopic imagery displayed on the screen and generate digital orthophotographs or digital elevation models. Also, only a few digital image processing systems incorporate expert systems or neural networks or fuzzy logic. Finally, systems of the future should provide detailed image lineage (genealogy) information about the processing applied to each image. The image lineage information (metadata) is indispensable when the products derived from the analysis of remotely sensed data are subjected to intense scrutiny as in environmental litigation.

Digital Image Processing Software

Some commonly-used digital image processing programs are listed in Table 3-5. The information provided in the table does not endorse any particular digital image processing software program. There is no "one-

TABLE 3–4 Image processing functions in quality digital image processing systems.

Preprocessing
o ***Radiometric correction*** of error introduced by the sensor system electronics/optics and/or environmental effects (includes relative image-to-image normalization and absolute radiometric correction of atmospheric attenuation)
o ***Geometric correction*** (image-to-image registration or image-to-map rectification)
Display and Enhancement
o Black-and-white computer display (8-bit)
o Color composite computer display (24-bit)
o Magnification (zooming), reduction, roaming
o Contrast manipulation (linear, nonlinear)
o Color space transformations (e.g., RGB to IHS)
o Image algebra (e.g., band ratioing, image differencing, NDVI, SAVI, Kauth-Thomas, EVI)
o Spatial filtering (e.g., low-pass, high-pass, bandpass)
o Edge enhancement (e.g., Kirsch, Laplacian, Sobel)
o Principal components analysis
o Texture transforms
o Frequency transformations (e.g., Fourier, Walsh)
o Digital elevation models (e.g., creation of TIN, interpolation via inverse distance weighting, or kriging, analytical hill shading, calculation of slope, aspect)
o Three-dimensional transformations (e.g., image draping over digital elevation models)
o Image animation (e.g., movies, change detection)
Information Extraction
o Pixel brightness value (BV_{ijk})
o Black-and-white or color density slice
o Transects (spatial and spectral)
o Univariate and multivariate statistical analysis
o Feature (band) selection (graphical and statistical)
o Supervised (e.g., minimum distance, maximum likelihood) and unsupervised classification (e.g., ISODATA)
o Geographic object-based image analysis (GEOBIA)
o Incorporation of ancillary data during classification
o Expert system image analysis including rule-based decision-tree classifiers and machine learning
o Neural network analysis and use of support vector machines
o Fuzzy logic classification
o Hyperspectral data analysis
o LiDAR data analysis
o RADAR data analysis
o Photogrammetric mapping (creation of orthoimages, extraction of digital elevation models and planimetric detail)
o Change detection
o Accuracy assessment (descriptive and analytical)
Image and Map Cartographic Composition
o Scaled Postscript level III output of images and maps
Integrated Geographic Information Systems
o Raster (image)-based GIS
o Vector (polygon)-based GIS (must allow polygon overlay)
Utilities
o Network (e.g., local area network, Internet)
o Image compression (single image, video)
o Import and export of various file formats
Metadata and Image/Map Lineage Documentation
o Metadata
o Complete image and GIS file processing history

size-fits-all" digital image processing software program that everyone should use. Before purchasing digital image processing software, analysts should carefully evaluate the functions and capabilities to see if they match current and potential image analysis requirements. In most cases, representatives from the digital image processing software companies will be pleased to demonstrate the capability of the software. Analysts should ask software representatives to demonstrate specific types of geospatial analysis that are of special interest.

Characteristics of the most commonly-used digital image processing programs are summarized in Table 3-5. Most of the software uses the Windows operating system. A robust remote sensing-oriented digital image processing system should be able to perform:

- radiometric and geometric correction;
- image enhancement (e.g., filtering, vegetation indices, principal components analysis);
- traditional image classification (e.g., unsupervised clustering, supervised maximum likelihood classification);
- decision-tree classification, neural network analysis, and machine-learning;
- geographic object-based image analysis (GEOBIA) (Blaschke et al., 2014);
- hyperspectral image analysis;
- LiDAR data analysis;
- RADAR image analysis;
- photogrammetric mapping; and
- change detection.

Basically, there are four types of digital image processing programs:

- those used to analyze any type of image (e.g., medical x-rays, terrestrial photographs, personal photographs) such as Adobe Photoshop. Most of these cannot perform atmospheric or geometric correction nor can they perform traditional supervised and/or unsupervised image classification. MATLAB, on the other hand, has a robust image processing toolbox that can be used to perform a variety of image enhancements, perform image restoration, image segmentation, and can geometrically rectify imagery (McAndrew, 2004; Gonzalez et al., 2009; MathWorks, 2014).
- those designed specifically for analyzing digital remote sensor data such as ERDAS Imagine, ENVI, PC Geomatica, IDRISI, *e*Cognition, Feature Analyst, LiDAR Analyst, LP360, SOCET, TNTmips, VIPER;
- sophisticated GIS software that has digital image processing capabilities such as ArcGIS and GRASS (Geographic Resources Analysis Support System); and
- computer-aided design (CAD) software with some image processing capability such as AUTOCAD.

Below are some observations about multispectral, hyperspectral, LiDAR, RADAR, photogrammetric mapping, change detection, and integrated digital image processing/GIS software. The observations relate to the information found in Table 3-5.

Multispectral Digital Image Processing Software

ERDAS Imagine, ENVI, PCI Geomatica, TNTmips, and IDRISI are heavily adopted digital image processing systems used for analysis of aircraft and satellite multispectral remote sensor data. All have radiometric and geometric preprocessing algorithms, a variety of image enhancement and analysis routines, and useful change detection modules. Some can process RADAR imagery.

Geographic Object-based Image Analysis (GEOBIA)

*e*Cognition, Feature Analyst, IDRISI, and ENVI have excellent geographic object-based image analysis (GEOBIA) programs. ERDAS and MATLAB also have image segmentation and processing capabilities.

Hyperspectral Digital Image Processing Software

ENVI and VIPER are rigorous hyperspectral image analysis programs. Both have sophisticated radiometric (atmospheric) correction capabilities and a diversity of algorithms to extract information from hyperspectral data. ENVI is a commercial product while VIPER is freeware available from the University of California at Santa Barbara. ACORN (Atmospheric Correction Now) is a commercial atmospheric correction program. ERDAS and IDRISI also have useful hyperspectral analysis capability. MultiSpec is a freeware program available from Purdue University with hyperspectral image processing functions.

LiDAR Digital Image Processing Software

SOCET GXP software has extensive enterprise level LiDAR 3D processing capability used primarily by intelligence gathering agencies (O'Neil-Dunne, 2012). LiDAR Analyst and QCoherent LP360 are sophisticated LiDAR 3D information extraction programs used extensively for civil engineering, intelligence-gathering and forestry applications. ArcGIS, GRASS, ERDAS, IDRISI, TNTmips, ER Mapper, and ENVI can create triangular-irregular network (TIN) or raster digital terrain models from LiDAR point cloud data (i.e., masspoints).

TABLE 3–5 Selected digital image processing software programs and their characteristics. The greater the capability, the greater the number of bullets •••••. The Internet address of the vendor or agency is provided.

Digital Image Processing Software	Operating System (Windows, Mac, Unix)	Preprocessing		Enhancement	Information Extraction						
		Rectification	Atmospheric		Traditional	Decision Tree	GOBIA	Hyperspectral	LiDAR	RADAR	Photogrammetric
ACORN® imspec.com	Windows		•••••								
ArcGIS® ArcMap® esri.com	Windows/ UNIX	••••		••••	••••						
AUTOCAD® autodesk.com	Windows/ Mac	•••••		••	••						
TNTmips® microimages.com	Windows	•••••	••••	•••••	•••••	•••••	••	••••	••	••	••
eCognition® ecognition.com	Windows	•••••	•••••	•••••	•••	•••••	•••••	•••••	••	••	•
ENVI® exelisvis.com	Windows/ UNIX	•••••	•••••	•••••	•••••	•••••	•••••	•••••	••••	••••	•••
ERDAS ER Mapper® intergraph.com	Windows/ UNIX	•••••	•••••	•••••	•••••	•••••	•••••	•••••	•••	••••	
ERDAS Imagine® intergraph.com	Windows/ UNIX	•••••	•••••	•••••	•••••	•••••	•••••	•••••	•••	•••••	•••
Feature Analyst® overwatch.com	Windows	•••••		•••••			•••••				
GRASS GIS® grass.osgeo.org	Windows/ Mac/UNIX	•••••	•••	•••••	•••••						
IDRISI® clarklabs.org	Windows/ UNIX	•••••	•••••	•••••	•••••	•••••	•••	•••	••	•••	
Leica Photogrammetry Suite (LPS)® intergraph.com	Windows/ UNIX	•••••	••	•••••							•••••
LiDAR Analyst® overwatch.com	Windows	*****							•••••		
LP360® qcoherent.com	Windows	•••••	•	••••					•••••		
MATLAB® Mathworks.com	Windows/ Mac/Linux	•••••	•	•••••	•••••	••	•••••				
MrSid® lizardtech.com	Windows/ Linux	image compression									
MultiSpec©® engineering.purdue.edu/~biehl/MultiSpec/	Windows/ Mac	•	•	•••	•••••			••••			
PCI Geomatica® pcigeomatics.com	Windows/ UNIX	•••••	•••••	•••••	•••••	•••••	•••••	•••••	•••••	•••••	••••
Photoshop® adobe.com	Windows/ Mac			•••••							

TABLE 3–5 (continued) Selected digital image processing software programs and their characteristics. The greater the capability, the greater the number of bullets •••••.

Digital Image Processing Software	Operating System (Windows, Mac, Unix)	Preprocessing			Information Extraction						
		Rectification	Atmospheric	Enhancement	Traditional	Decision Tree	OBIA	Hyperspectral	LiDAR	RADAR	Photogrammetric
R® www.r-project.org	Windows/ Mac			•••••	••••	•••••	•••				
SOCET GXP® socetgxp.com	Windows/ UNIX/Linux	•••••	•••••	•••••	•••••	•••••	•••••	•••••	•••••	•••••	•••••
VIPER® vipertool.org	Windows	•••••	•••••	••	••			•••••			

RADAR Digital Image Processing Software

Only a few of the commercially-available digital image processing programs can process RADAR (single polarization) or Polarimetric RADAR (multiple polarization) data. ERDAS Imagine, ENVI, PCI Geomatica, and IDRISI have RADAR processing modules.

Photogrammetric Mapping Software

Photogrammetric mapping involves the processing of stereoscopic digital aerial photography to extract planimetric information (e.g., building footprints, road centerlines, fences, tree footprints, drainage networks) and/or three-dimensional topographic information. SOCET software by BAE Systems, Inc., is a quality soft-copy photogrammetry program. However, few outside of intelligence-gathering agencies have access to its capabilities (O'Neil-Dunne, 2012). Intergraph's Leica Photogrammetry Suite (LPS) has a user-friendly interface and extensive photogrammetric mapping capabilities. ERDAS, ENVI, IDRISI, TNTmips, and ER Mapper have some soft-copy photogrammetric processing functions, most notably those necessary to create orthophotography, and digital terrain models.

Change Detection

Some software such as ERDAS (e.g., DeltaQue), IDRISI (e.g., Land Change Modeler for ArcGIS), ENVI (e.g., SPEAR), and Esri (e.g., Change Matters; Green, 2011) have change detection programs or wizards for analyzing multiple date aircraft and/or satellite data (Jensen et al., 2012). Pictometry Analytics (2014) markets ChangeFindr software, which works with both vertical and oblique digital aerial photography.

Integration of Digital Image Processing and GIS Functions

ArcGIS is the most sophisticated vector-based geographic information system. Remote sensing data and remote sensing-derived thematic information is easily incorporated and analyzed in ArcGIS ArcMap. IDRISI and GRASS are powerful raster-based geographic information systems that easily incorporate vector information for GIS modeling. ERDAS and ENVI can incorporate vector information in their various GIS analysis operations. ERDAS Apollo is a comprehensive digital image processing, photogrammetry, and GIS program (du Plessis, 2012). TNTmips is an integrated GIS, image processing, CAD, desktop cartography, and geospatial database management program.

Cost

Companies, public agencies, and academic institutions have limited financial resources that must be used carefully. Therefore, the cost of commercial digital image processing software is a serious consideration. Digital image processing software is expensive. This may be the result of the still closed-source nature of most digital image processing software. The price for a single digital image processing software license is highest for commercial users, somewhat lower for public agencies, and much lower for academic institutions.

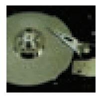
Open-Source Digital Image Processing Software

If software cost is a concern, open-source digital image processing software, such as GRASS or MultiSpec, may be the best solution.

Open-Source Statistical Analysis Software that can be used for Digital Image Processing

In addition to commercial SAS and SPSS statistical analysis software packages, advanced digital image processing can be performed using **R**—an open-source programming language available at http://www.r-project.org/. **R** is widely used by statisticians and data miners to develop statistical analysis software. The types of processes that can be applied to remote sensing images are available in various user-submitted **R** packages (Venables et al., 2012). Image analysts often use packages from the *Machine Learning & Statistical Learning* group to perform neural network analysis, decision-

tree analysis, CART, random forests, C5.0, support vector machine (SVM) analysis, and others discussed in Chapter 10.



Digital Image Processing and the National Spatial Data Infrastructure

Laypersons and scientists who use remote sensing data or share products derived from remotely sensed data should be aware of very specific spatial data standards developed by the Federal Geographic Data Committee (FGDC). The FGDC is an interagency committee of representatives from the Executive Office of the President, the Cabinet, and independent agencies. The FGDC is developing the National Spatial Data Infrastructure (NSDI) in cooperation with organizations from state, local and tribal governments, academics, and the private sector. The NSDI encompasses policies, standards, and procedures for organizations to cooperatively produce and share geographic data (FGDC NSDI, 2014).

The American National Standards Institute's (ANSI) Spatial Data Transfer Standard (SDTS) is a mechanism for archiving and transferring spatial data (including metadata) between dissimilar computer systems. The SDTS specifies exchange constructs, such as format, structure, and content, for spatially referenced vector and raster (including gridded) data. Actual use of SDTS to transfer spatial data is carried out through its profiles. The FGDC Raster Profile standard is of particular interest because it provides specifications for transferring spatial datasets in which features or images are represented in raster or gridded form, such as digital elevation models, digital orthophoto quarter quads (DOQQ), and digital satellite imagery (FGDC SDTS, 2014).



References

Amdahl, G., 1967, "Validity of the Single Processor Approach to Achieving Large-Scale Computing Capabilities," *AFIPS Conference Proceedings,* 30:483–485.

Blaschke, T., and 10 co-authors, 2014, "Geographic Object-Based Image Analysis – Towards A New Paradigm," *ISPRS Journal of Photogrammetry & Remote Sensing,* 87:180–191.

Canty, M. J., 2014, *Image Analysis, Classification and Change Detection in Remote Sensing: With Algorithms for ENVI/IDL and Python*, (3rd Ed.), Boca Raton: CRC Press, 576 p.

du Plessis, S., 2012, "Photogrammetry, LiDAR, Remote Sensing and GIS Together at Last," *LiDAR*, 2(5):24–28.

FGDC SDTS, 2014, *Spatial Data Transfer Standard*, Washington: USGS, http://mcmcweb.er.usgs.gov/sdts/.

FGDC NSDI, 2014, *The National Spatial Data Infrastructure*, Washington: Federal Geographic Data Committee, https://www.fgdc.gov/nsdi/nsdi.html.

Gonzalez, R. C., Woods, R. E., and S. L. Eddins, 2009, *Digital Image Processing Using MATLAB*, 2nd Ed., New York: Gatesmark Publishing, 50 p.

Green, K., 2011, "Change Matters," *Photogrammetric Engineering & Remote Sensing*, 77(4):305–309.

Intel, 2014, *The Evolution of a Revolution*, Intel, Inc., http://www.intel.com/pressroom/kits/quickref.htm.

Jensen, J. R., 2007, *Remote Sensing of the Environment: An Earth Resource Perspective,* Boston: Pearson, Inc., 554 p.

Jensen, J. R. and R. R. Jensen, 2013, *Introductory Geography Information Systems*, Boston: Pearson, Inc., 400 p.

Jensen, J. R., Guptill, S., and D. Cowen, 2012, *Change Detection Technology Evaluation Report,* Task T007, Washington: U.S. Census Bureau, August 20, 2012, 274 p.

Kouyoumjia, V., 2010, "The New Age of Cloud Computing and GIS," *ArcWatch*, January, 2010 (http://www.Esri.com/news/arcwatch/0110/feature.html).

MacNeil, T., 2004, "Don't be Misled By MIPS," *IBM Systems Magazine*, www.ibmsystemsmag.com/mainframe/tipstechniques/systemsmanagement/Don-t-Be-Misled-By-MIPS/.

MathWorks, 2014, *MATLAB: The Language of Technical Computing*, New York, Mathlab, Inc., http://www.mathworks.com/products/matlab/

McAndrew, A., 2004, *An Introduction to Digital Image Processing with MATLAB*, Victoria Univ. of Technology: School of Computer Science & Mathematics, http://visl.technion.ac.il/labs/anat/An%20Introduction%20To% 20 Digital%20Image%20Processing%20With%20Matlab.pdf.

O'Neil-Dunne, J., 2012, "Review of SOCET GXP," *LiDAR*, 2(5):54–59.

Pictometry Analytics, 2014, *ChangeFindr*, http://www.eagleview.com/Products/ImageSolutionsAnalytics/PictometryAnalyticsDeployment.aspx#ChangeFinder.

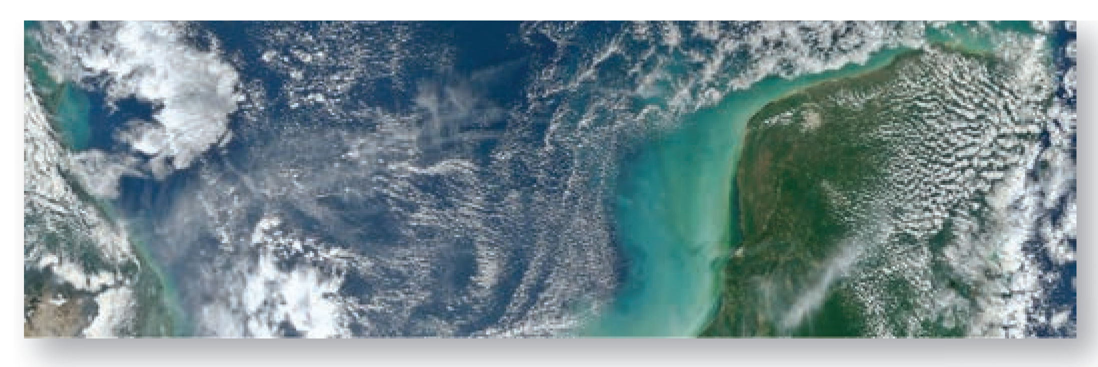
Rothenberg, J., 1995, "Ensuring the Longevity of Digital Documents," *Scientific American*, 272:42–47.

Sorokine, A, 2007, "Implementation of a Parallel High-performance Visualization Technique in GRASS GIS," *Computers & Geosciences*, 33(5):685–695.

Venables, W. N., Smith, D. M. and the R Core Team, 2012, *An Introduction to R: Notes on R - Programming Environment for Data Analysis and Graphics*, Version 2.14.2 (2012–10-26), 109 p., http://www.r-project.org/.

Yang, C. and Q. Huang, 2013, *Spatial Cloud Computing: A Practical Approach*, Boca Raton: CRC Press, 357 p.

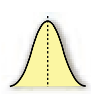
4 Image Quality Assessment and Statistical Evaluation



Source: NOAA

Many remote sensing datasets contain high-quality, accurate data. Unfortunately, sometimes error (or noise) is introduced into the remote sensor data by a) the environment (e.g., atmospheric scattering), b) random or systematic malfunction of the remote sensing system (e.g., an uncalibrated detector creates striping), or c) improper airborne or ground processing of the remote sensor data prior to actual data analysis (e.g., an inaccurate analog-to-digital conversion). Therefore, the person responsible for analyzing the digital remote sensor data should first assess its quality and statistical characteristics. This is normally accomplished by:

- looking at the frequency of occurrence of individual brightness values in the image displayed in histogram format,
- viewing on a computer monitor individual pixel brightness values at specific locations or within a geographic area,
- computing fundamental univariate descriptive statistics to determine if there are unusual anomalies in the image data, and
- computing multivariate statistics to determine the amount of between-band correlation (e.g., to identify redundancy).



Overview

This chapter first reviews basic elements of statistical sampling theory. It then introduces the histogram and its significance to digital image processing of remote sensor data. Various methods of viewing individual pixel values and geographic areas of individual pixel values are then presented. Algorithms for the computation of univariate and multivariate statistics are introduced, including the identification of the minimum and maximum value for each band of imagery, the range, mean, standard deviation, and between-band covariance and correlation. Finally, geostatistical analysis is introduced that can be of value for obtaining information about spatial autocorrelation in imagery and when performing spatial interpolation.



Image Processing Mathematical Notation

The following notation is used to describe the mathematical operations applied to the digital remote sensor data:

i = a row (or line) in the imagery

j = a column (or sample) in the imagery

k = a band of imagery

l = another band of imagery

n = total number of picture elements (pixels) in an array

BV_{ijk} = brightness value at row i, column j, and band k

BV_{ik} = ith brightness value in band k

BV_{il} = ith brightness value in band l

min_k = minimum brightness value of band k

max_k = maximum brightness value of band k

$range_k$ = range of brightness values in band k

$quant_k$ = quantization level of band k (e.g., 2^8 = 0 to 255; 2^{12} = 0 to 4,095)

μ_k = mean of band k

var_k = variance of band k

s_k = standard deviation of band k

$skewness_k$ = skewness of a band k distribution

$kurtosis_k$ = kurtosis of a band k distribution

cov_{kl} = covariance between pixel values in two bands, k and l

r_{kl} = correlation between pixel values in two bands, k and l

X_c = measurement vector for class c composed of brightness values (BV_{ijk}) from row i, column j, and band k

M_c = mean vector for class c

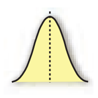
M_d = mean vector for class d

⊠$_{ck}$ = mean value of the data in class c, band k

s_{ck} = standard deviation of the data in class c, band k

v_{ckl} = covariance matrix of class c for bands k through l; shown as V_c

v_{dkl} = covariance matrix of class d for bands k through l; shown as V_d



Sampling Theory

Digital image processing is performed on only a sample of all available remote sensing information. Therefore, it is useful to review several fundamental aspects of elementary statistical sampling theory. A **population** is an infinite or finite set of elements. An infinite population would be all possible images that might be acquired of the entire Earth in 2013. All Pleiades images of Charleston, SC, obtained in 2013 would be a finite population.

Types of Sampling

A **sample** is a subset of the elements taken from a population used to make inferences about certain characteristics of the population. For example, we might decide to analyze a November 1, 2013, Pleiades image of Charleston, SC. If observations with certain characteristics are systematically excluded from the sample either deliberately or inadvertently (such as selecting images obtained only in the fall of the year), it is a **biased sample**. The difference between the true value of a population characteristic and the value of that characteristic inferred from a sample is the **sampling error**.

The goal of sampling is to collect an unbiased representative sample of the population (Jensen and Shumway, 2010). A **random sample** with replacement occurs when every observation has an equal chance of being selected. This is one of the most simple sampling schemes, and it usually ensures that there is no bias in the sample. An example of a simple random sample used to assess the accuracy of a remote sensing-derived thematic map is presented in Chapter 13.

A **systematic sample** can be implemented according to a predetermined system (e.g., collect data every 100 m in the x- and y-directions). In effect, most remote sensing data are collected systematically (e.g., Landsat 8 systematically collects panchromatic imagery every 15 ⊠ 15 m). An example of a systematic sample used to assess the accuracy of a remote sensing-derived thematic map is presented in Chapter 13.

In some cases, a stratified sample is most appropriate. A **stratified random sample** could be implemented when the analyst knows that the study area contains different sub-populations and he or she makes an effort to sample within each sub-population (i.e., strata). This helps ensure that all of the variation present in the spatial dataset is accounted for (Jensen and Shumway, 2010). An example of a stratified random sample used to assess the accuracy of a remote sensing-derived thematic map is presented in Chapter 13.

Large samples drawn randomly from natural populations usually produce a symmetrical frequency distribution such as that shown in Figure 4-1a. Most values are clustered around a central value, and the frequency of occurrence declines away from this central point. A graph of the distribution appears bell-shaped and is called a **normal distribution**. Many statistical tests used in the analysis of remotely sensed data assume that the reflectance, emittance, or back-scattered values recorded in a scene are normally distributed. Unfortunately,

Histograms of Symmetric and Skewed Distributions

Symmetric

Skewed

Mean
Median
Mode

Frequency

a.

Normal Distribution

Mode
Median
Mean

Frequency

d.

Negatively Skewed Distribution

Bimodal

Mean
Median

Frequency

b.

Multimodal Distribution

Mode
Median
Mean

Frequency

e.

Positively Skewed Distribution

Mean
Median

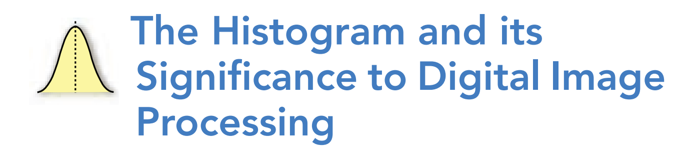
Frequency

c.

Uniform Distribution,
no mode exists

FIGURE 4–1 Relative position of measures of central tendency for commonly encountered frequency distributions. a) Normally distributed data. b) Bi-modal distribution. c) Uniform distribution. d) Negatively skewed distribution. e) Positively skewed distribution.

remotely sensed data may not be normally distributed, and the analyst must be careful to identify such conditions. In such instances, nonparametric statistical theory may be preferred.



The Histogram and its Significance to Digital Image Processing

The histogram is a useful graphic representation of the information content of a single band of remotely sensed data (Papp, 2010). Histograms for each band of imagery are often displayed and analyzed in many remote sensing investigations because they provide the analyst with an appreciation of the quality of the original data (e.g., whether it is low in contrast, high in contrast, or multimodal in nature). In fact, many analysts routinely provide before (original) and after histograms of the imagery to document the effects of applying an image enhancement technique (Gonzalez and Woods, 2007; Pratt, 2007; Russ, 2011). It is instructive to review how a histogram of a single band of imagery, k, composed of i rows and j columns with a brightness value BV_{ijk} at each pixel location is constructed.

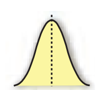
Individual bands of remote sensor data are typically quantized (digitally recorded) with brightness values ranging from 2^8 to 2^{12} (if $quant_k = 2^8$ then brightness values range from 0 to 255; 2^9 = values from 0 to 511; 2^{10} = values from 0 to 1,023; 2^{11} = values from 0 to

2,047; and 2^{12} = values from 0 to 4,095). Much of the remote sensor data are quantized to 8 bits, with values ranging from 0 to 255 (e.g., Landsat 5 Thematic Mapper and SPOT HRV data). Some sensor systems such as GeoEye-1 and MODIS obtain data with 11 bits of precision. The greater the quantization, the higher the probability that more subtle spectral reflectance (or emission) characteristics may be extracted from the imagery (Bossler et al., 2010).

Tabulating the frequency of occurrence of each brightness value within the image provides statistical information that can be displayed graphically in a histogram. The range of quantized values of a band of imagery, $quant_k$, is provided on the abscissa (*x*-axis), while the frequency of occurrence of each of these values is displayed on the ordinate (*y*-axis) (Figure 4-1a). For example, consider the characteristics of a Landsat 4 Thematic Mapper image of Charleston, SC, collected on November 9, 1982, displayed as a color-infrared color composite in Figure 4-2a. Band 4 (near-infrared) is displayed in Figure 4-2b. A histogram of the band 4 image data is displayed in Figure 4-2c. The peaks in the histogram correspond to dominant types of land cover in the image, including: ‘a’ open water pixels, ‘b’ coastal wetland, and ‘c’ upland. Note how the Landsat Thematic Mapper band 4 data are compressed into only the lower one-third of the 0 to 255 range, suggesting that the data are relatively low in contrast. If the original Landsat Thematic Mapper band 4 brightness values were displayed on a monitor screen or on the printed page they would be relatively dark and difficult to interpret. Therefore, in order to see the wealth of spectral information in the scene, the original band 4 brightness values were contrast stretched (Figure 4-2b). Contrast stretching principles are discussed in Chapter 8.

Histograms are useful for evaluating the quality of optical daytime multispectral data and many other types of remote sensor data. For example, consider the histogram of predawn thermal infrared (8.5 to 13.5 μm) imagery of a thermal plume in the Savannah River shown in Figure 4-3. The thermal plume entered the Savannah River via Four Mile Creek, which carried hot water used to cool industrial activities. The peaks in this histogram are associated with ‘a’ the relatively cool temperature of the Savannah River swamp on each side of the river’s natural levee, ‘b’ the slightly warmer temperature (12°C) of the Savannah River upstream and west of the plume, and ‘c’ the relatively hot thermal plume that progressed through the swamp and then entered the Savannah River. The north to south flow of the Savannah River caused the plume to hug the eastern bank and dissipate as it progressed away from the mouth of the creek.

When an unusually large number of pixels have the same brightness value, the traditional histogram display might not be the best way to communicate the information content of the remote sensor data. When this occurs, it might be useful to scale the frequency of occurrence (*y*-axis) according to the relative percentage of pixels within the image at each brightness level along the *x*-axis.



Metadata

Metadata is data or information about data. Most quality digital image processing systems read, collect, and store metadata about a particular image or subimage. It is important that the image analyst has access to this metadata. In the most fundamental instance, metadata might include: the file name, date of last modification, level of quantization (e.g, 8 bits), number of rows and columns, number of bands, univariate statistics (minimum, maximum, mean, median, mode, standard deviation), georeferencing performed (if any), and pixel dimension (e.g., 5 × 5 m). Utility programs within the digital image processing system routinely provide this information. For example, fundamental metadata about the Charleston, SC, Landsat 4 Thematic Mapper subscene and the Savannah River thermal infrared subscene are found in Figures 4-2 and 4-3, respectively. Neither of these images has been subjected to geometric rectification as yet. Therefore, there is no metadata concerning the projection, spheroid, or datum. This information is recorded as metadata later on in the process when geometric rectification takes place.

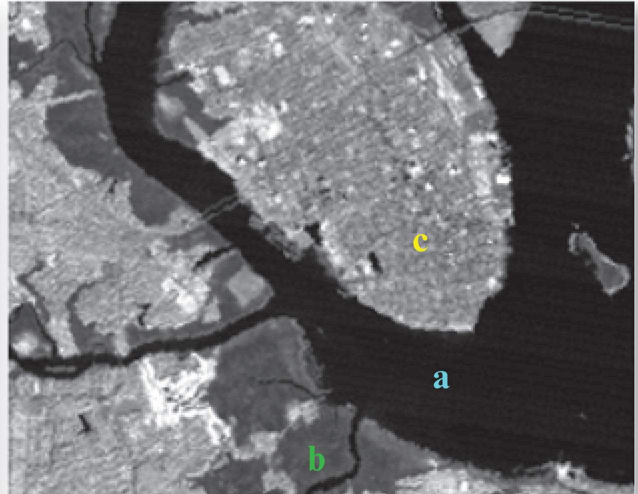
An ideal remote sensing metadata system would keep track of every type of processing applied to each digital image (Lanter, 1991). This ‘image genealogy’ or ‘lineage’ information can be very valuable when the remote sensor data are subjected to intense scrutiny (e.g., in a public forum), used in litigation, if you want to repeat a procedure, or co-register two datasets.

The Federal Geographic Data Committee (FGDC) has set up rigorous image metadata standards as part of the National Spatial Data Infrastructure (NSDI) (FGDC, 2014). All federal agencies that provide remote sensor data to the public are required to use the established metadata standards. For example, when someone purchases a 1 × 1 m digital orthophotoquad from the U.S. Geological Survey produced from 1:40,000-scale National Aerial Photography Program (NAPP) data, it comes with detailed metadata. Image analysts should always consult the image metadata and incorporate the information in the digital image processing that takes place. Almost all commercial remote

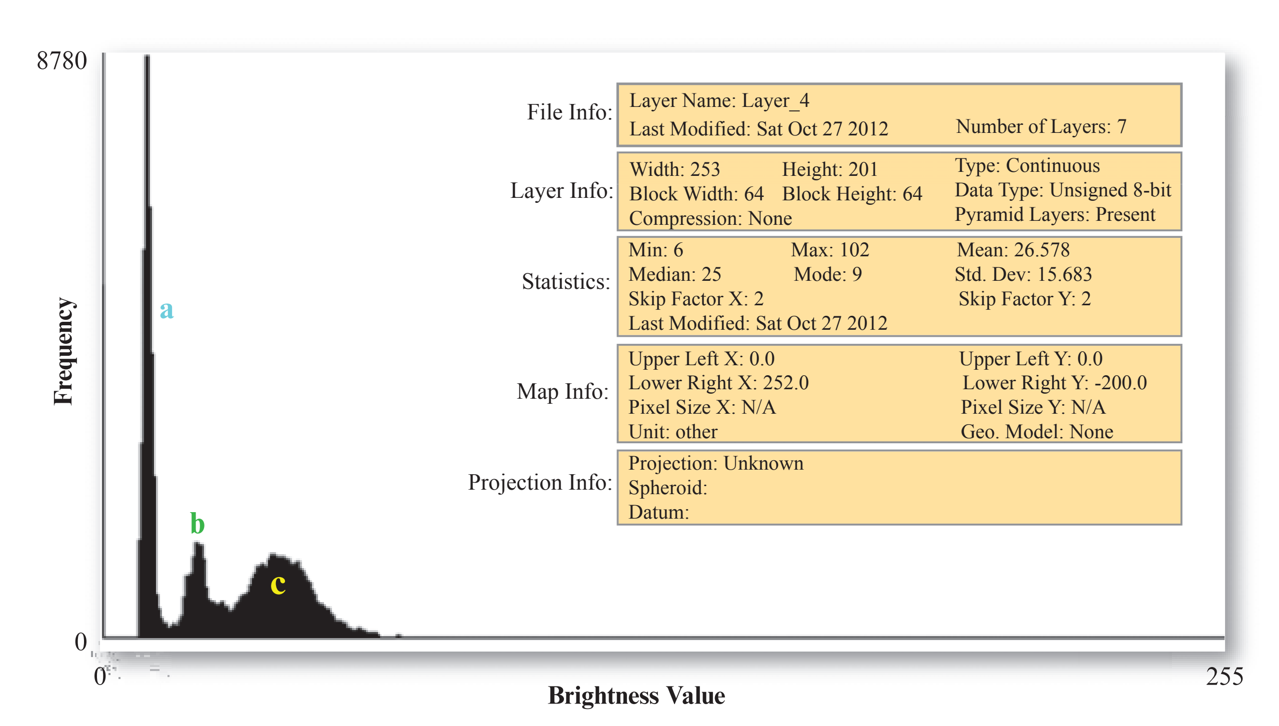
Landsat 4 Thematic Mapper Image of Charleston, SC, Collected on November 9, 1982



a. Color composite of bands 4 (near-infrared), 3 (red), and 2 (green).



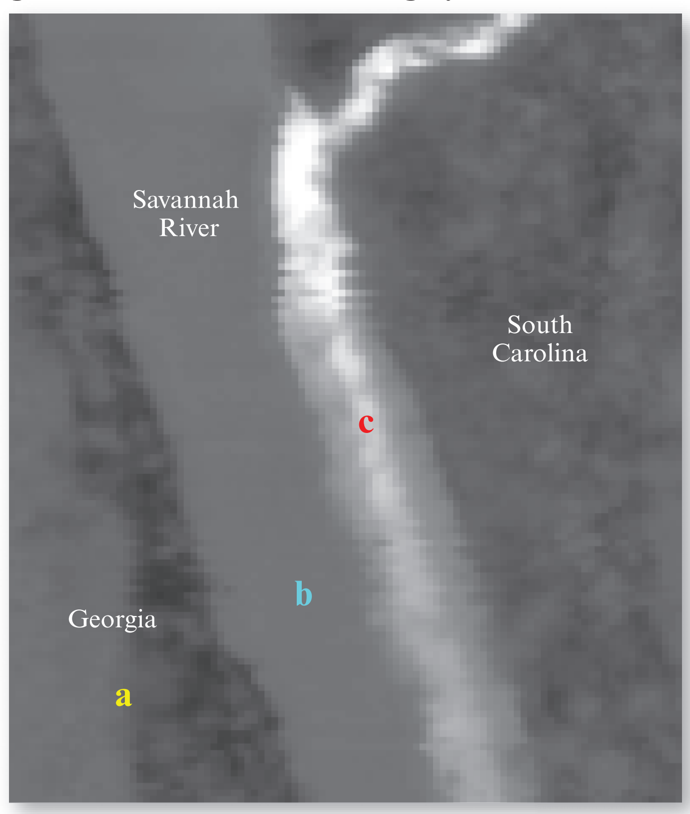
b. Band 4 (contrast stretched).



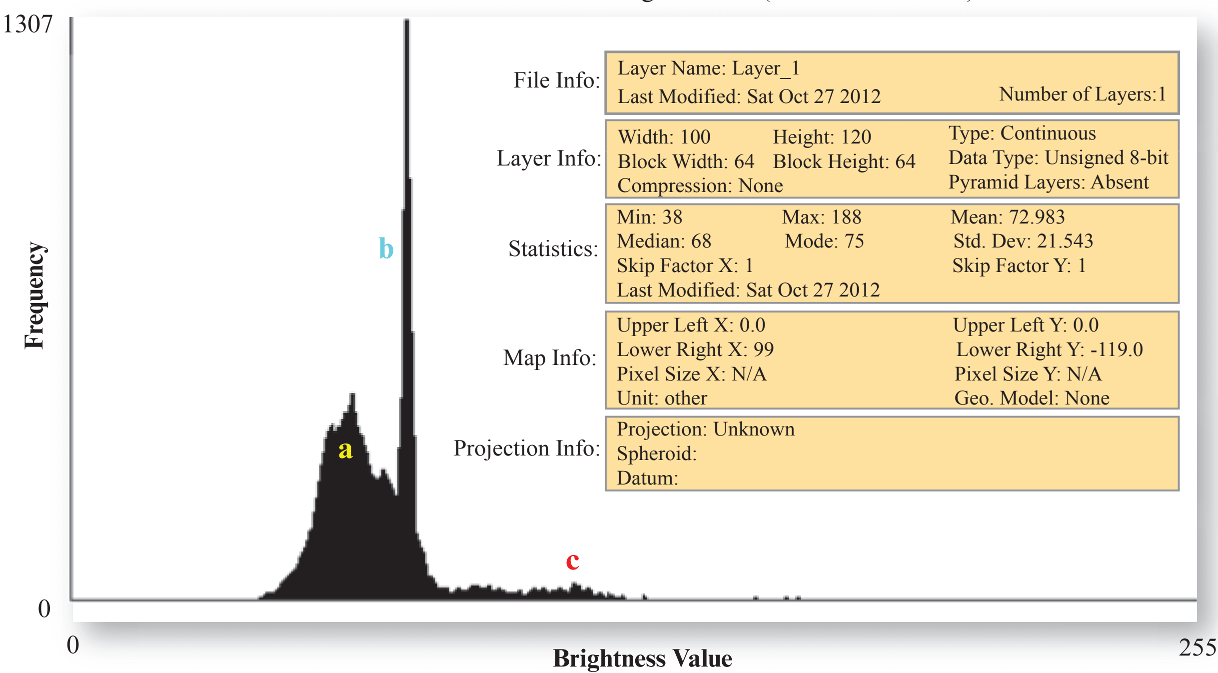
c. Histogram of the original Landsat Thematic Mapper band 4 image.

FIGURE 4–2 a) Landsat 4 Thematic Mapper (TM) image of Charleston, SC, obtained on November 9, 1982, displayed as a color-infrared color composite using bands 4 (near-infrared), 3 (red), and 2 (green). b) Display of the band 4 (near-infrared) TM image. c) A multimodal histogram of the brightness values of the band 4 image. Peaks in the histogram correspond to dominant types of land cover in the image, including 'a' open water pixels, 'b' coastal wetland consisting primarily of smooth cordgrass (*Spartina alterniflora*), and 'c' upland. The inset provides basic metadata about the characteristics of the Landsat Thematic Mapper band 4 subscene. Original imagery courtesy of NASA.

Histogram of Thermal Infrared Imagery of the Savannah River

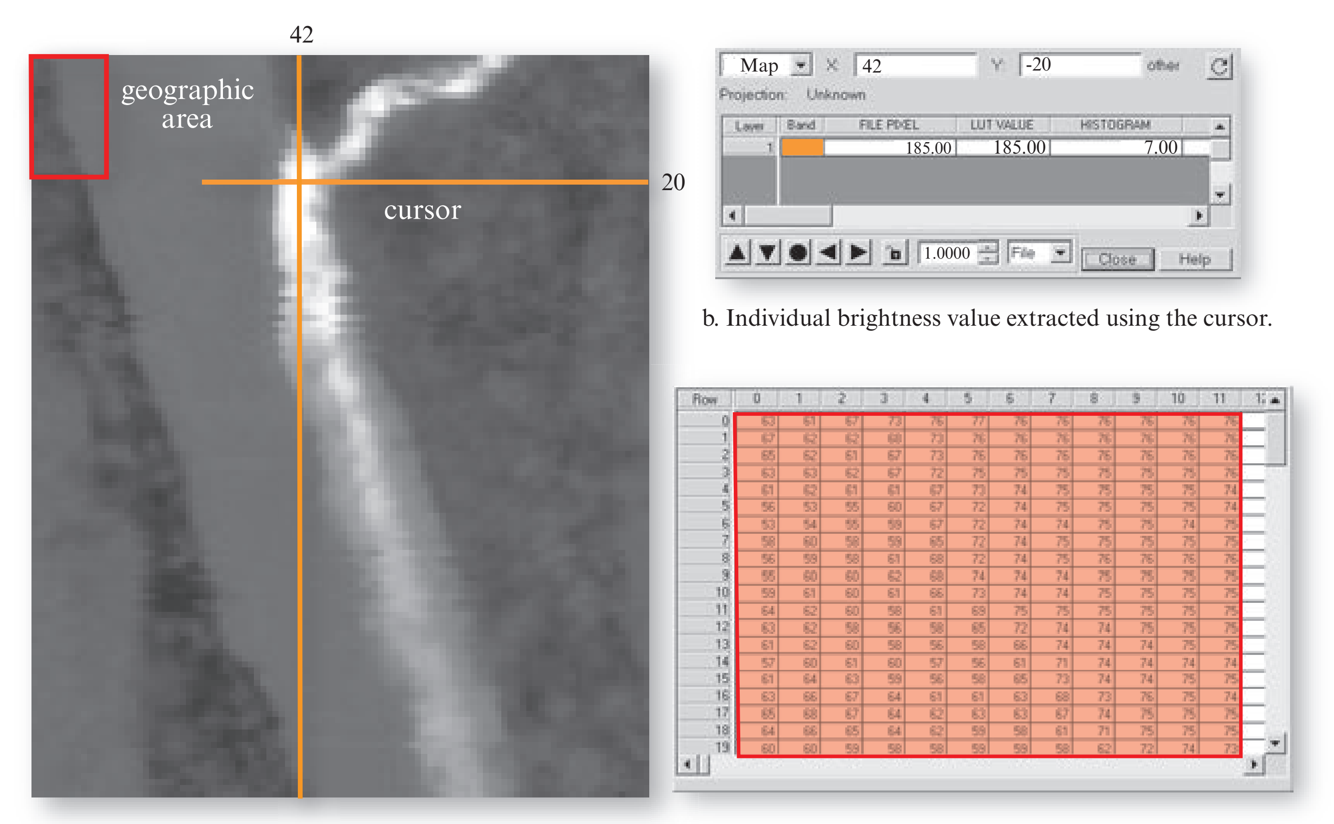


a. Nighttime thermal infrared image of the Savannah River on the South Carolina and Georgia border (contrast stretched).



b. Histogram of the original thermal infrared data.

FIGURE 4–3 a) Thermal infrared image of a thermal effluent plume in the Savannah River obtained on March 28, 1981, at 4:00 a.m. b) A multimodal histogram of the brightness values in the predawn thermal-infrared image. Peaks in the histogram are associated with 'a' the relatively cool temperature of the surrounding Savannah River swamp, 'b' the slightly warmer temperature (12°C) of the Savannah River, and 'c' the relatively hot thermal plume. The inset provides metadata about the characteristics of the thermal infrared image subscene.



a. Nighttime thermal infrared image of the Savannah River.

b. Individual brightness value extracted using the cursor.

c. Brightness values in a geographic area for an individual band.

FIGURE 4–4 a) Thermal infrared image of the Savannah River with a cursor located at column 42 (*x*) and row 20 (*y*). b) Spectral information located at column 42 and row 20. c) Brightness values found in the first 12 columns and first 20 rows of the thermal infrared image displayed in a matrix format.

sensing data providers are adopting FGDC metadata standards.

The histogram and metadata information help the analyst understand the content of remotely sensed data and judge its quality. Sometimes, however, it is very useful to look at individual brightness values at specific locations in the imagery.



Viewing Individual Pixel Values at Specific Locations or within a Geographic Area

Viewing individual pixel brightness values in a remotely sensed image is one of the most useful methods for assessing the quality and information content of the data. Virtually all digital image processing systems allow the analyst to:

- use a mouse-controlled cursor (cross-hair) to identify a specific geographic location in the image (at a particular row and column or geographic *x*,*y* coordinate) and display its brightness value in *n* bands, and
- display the individual brightness values of an individual band in a matrix (raster) format.

Cursor Evaluation of Individual Pixel Brightness Values

Most people know how to control a cursor using a mouse and navigate to a desired location. For example, a cursor has been placed in the heart of the Savannah River thermal plume at row 42 (*x*) and column 20 (*y*) in Figure 4-4a. The numeric information for this location in the thermal infrared image, summarized in Figure 4-4b, includes the original brightness value (185), the color look-up table value (discussed in Chapter 5), and the number of pixels in the histogram with a value of 135 (i.e., 7). Note the directional keys that can be used to navigate to nearby pixel locations if desired.

Two- and Three-Dimensional Evaluation of Pixel Brightness Values within a Geographic Area

It can become quite tedious using the cursor to evaluate the individual pixel brightness values throughout even a small geographic area. In such circumstances, it is useful to identify with the cursor a geographic area (e.g., a rectangle) and display all the pixel values in it for a specific band or bands. For example, Figure 4-4c displays in a matrix format the brightness values found in the first 20 rows and 12 columns of the thermal infrared image highlighted in the upper-left in Figure 4-4a. Brightness values >70 represent Savannah River pixels as well as the pixels found at the interface of the water and land. In Chapter 8 (Image Enhancement) we will perform a more rigorous examination of the thermal infrared data and select only those brightness values >74 as being relatively pure water pixels.

The display of individual brightness values in a matrix (raster) format is informative, but it does not convey any visual representation of the magnitude of the data within the area. Therefore, it is often useful early in the exploratory phases of a digital image processing project to extrude the value of the individual brightness values within a geographic area to create a pseudo three-dimensional display. For example, a wire-frame pseudo three-dimensional display of the individual brightness values over the entire thermal infrared image is shown in Figure 4-5b. While the wire-frame display is informative, it is usually more visually effective to drape the actual grayscale image of the individual band (or bands if a color composite is being used) over the exaggerated three-dimensional representation of the brightness values found within the specific geographic area. For example, a pseudo three-dimensional display of the thermal infrared image viewed from two different vantage points is shown in Figure 4-5c.

The examination of histograms and the extraction of individual brightness values using a cursor or geographic area analysis methods are quite useful. However, they provide no statistical information about the remote sensor data. The following sections describe how fundamental univariate and multivariate statistics are computed for the remotely sensed imagery.



Univariate Descriptive Image Statistics

Most digital image processing systems can perform robust univariate and multivariate statistical analyses of single- and multiple-band remote sensor data. For example, image analysts have at their disposal statistical measures of central tendency and measures of dispersion that can be extracted from the imagery.

Measure of Central Tendency in Remote Sensor Data

The **mode** (e.g., see Figure 4-1a) is the value that occurs most frequently in a distribution and is usually the highest point on the curve (histogram). It is common, however, to encounter more than one mode in a remote sensing dataset, such as that shown in Figure 4-1b. The histogram of the Landsat TM band 4 image of Charleston, SC (Figure 4-2c) and the predawn thermal infrared image of the Savannah River (Figure 4-3b) have multiple modes. They are referred to as being multimodal histograms (Figure 4-1b). They are nonsymmetrical (skewed) distributions.

The **median** is the value midway in the frequency distribution (e.g., see Figure 4-1a). One-half of the area below the distribution curve is to the right of the median, and one-half is to the left. The **mean** (μ) is the arithmetic average and is defined as the sum of all brightness value observations divided by the number of observations. It is the most commonly used measure of central tendency. The mean of a single band of imagery, μ_k, composed of n brightness values (BV_{ik}) is computed using the formula:

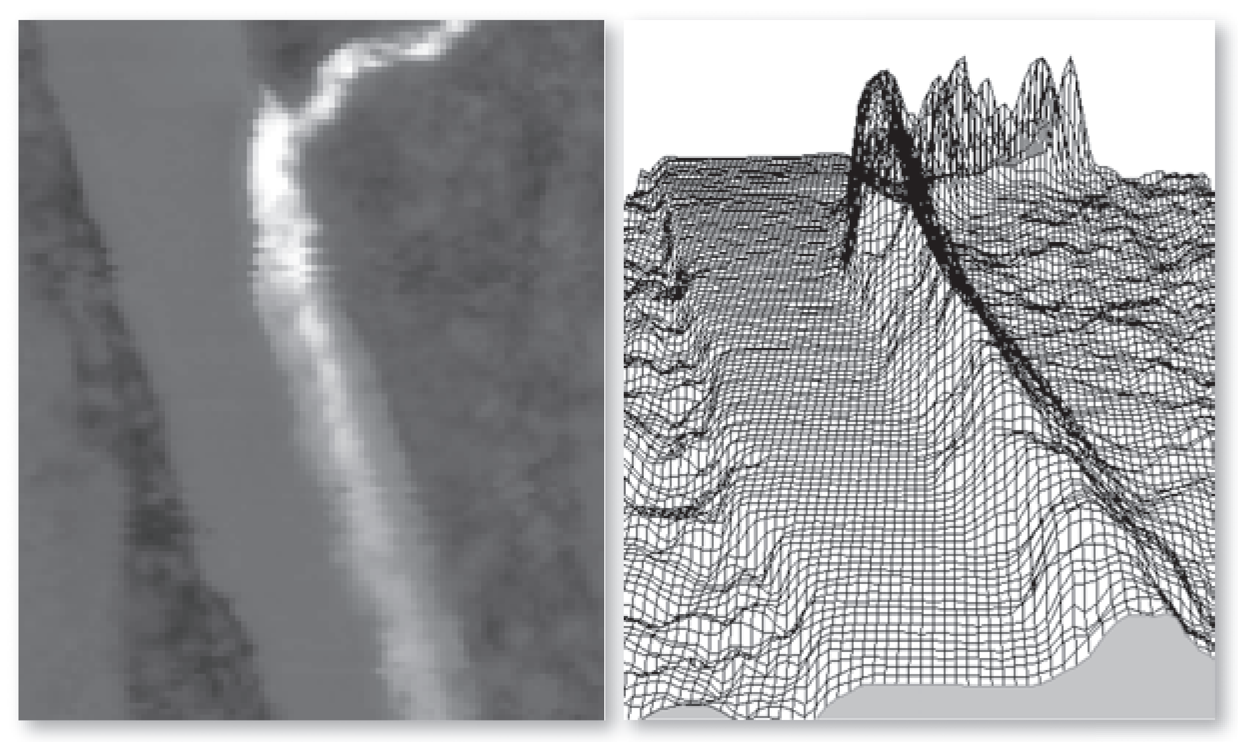
$$\mu_k = \frac{\sum_{i=1}^{n} BV_{ik}}{n} \,. \qquad (4.1)$$

The sample mean, μ_k, is an unbiased estimate of the population mean. For symmetrical distributions the sample mean tends to be closer to the population mean than any other unbiased estimate (such as the median or mode). Unfortunately, the sample mean is a poor measure of central tendency when the set of observations is skewed or contains an extreme value (outlier). As the peak (mode) becomes more extremely located to the right or left of the mean, the frequency distribution is said to be **skewed**. A frequency distribution curve (histogram) is said to be skewed in the direction of the longer tail. Therefore, if a peak (mode) falls to the right of the mean, the frequency distribution is negatively skewed. If the peak falls to the left of the mean, the frequency distribution is positively skewed. Examples of positively and negatively skewed distributions are shown in Figures 4-1d and e.

Measures of Dispersion

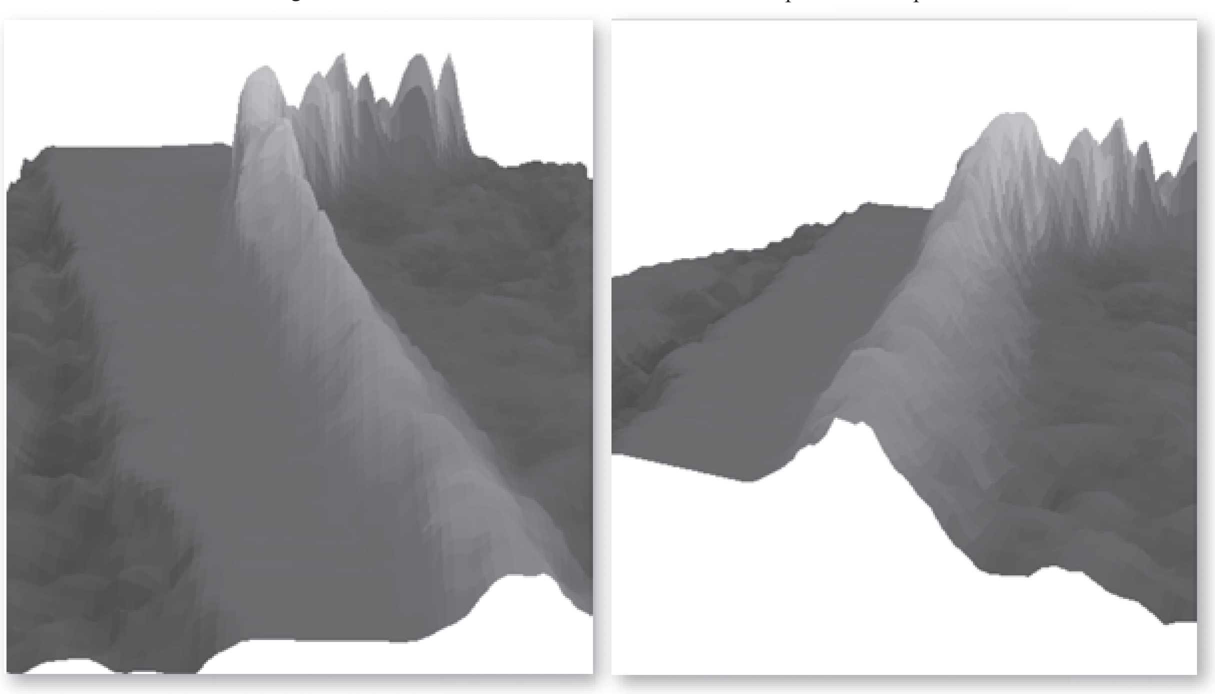
Measures of the dispersion about the mean of a distribution also provide valuable information about the image. For example, the **range** of a band of imagery

Thermal Infrared Image Draped over Three-Dimensional Representation of the Data



a. Thermal infrared image of the Savannah River.

b. Wire-frame pseudo 3D representation of the data.



c. Thermal infrared data draped over the 3D representation (5× exaggeration).

FIGURE 4–5 Two- and three-dimensional views of the Savannah River thermal infrared data shown in Figure 4-3a.

($range_k$) is computed as the difference between the maximum (max_k) and minimum (min_k) values:

$$range_k = max_k - min_k . \qquad (4.2)$$

Unfortunately, when the minimum or maximum values are extreme or unusual observations (i.e., possibly **data blunders**), the range could be a misleading measure of dispersion. Such extreme values are common because the remote sensor data are often collected by detector

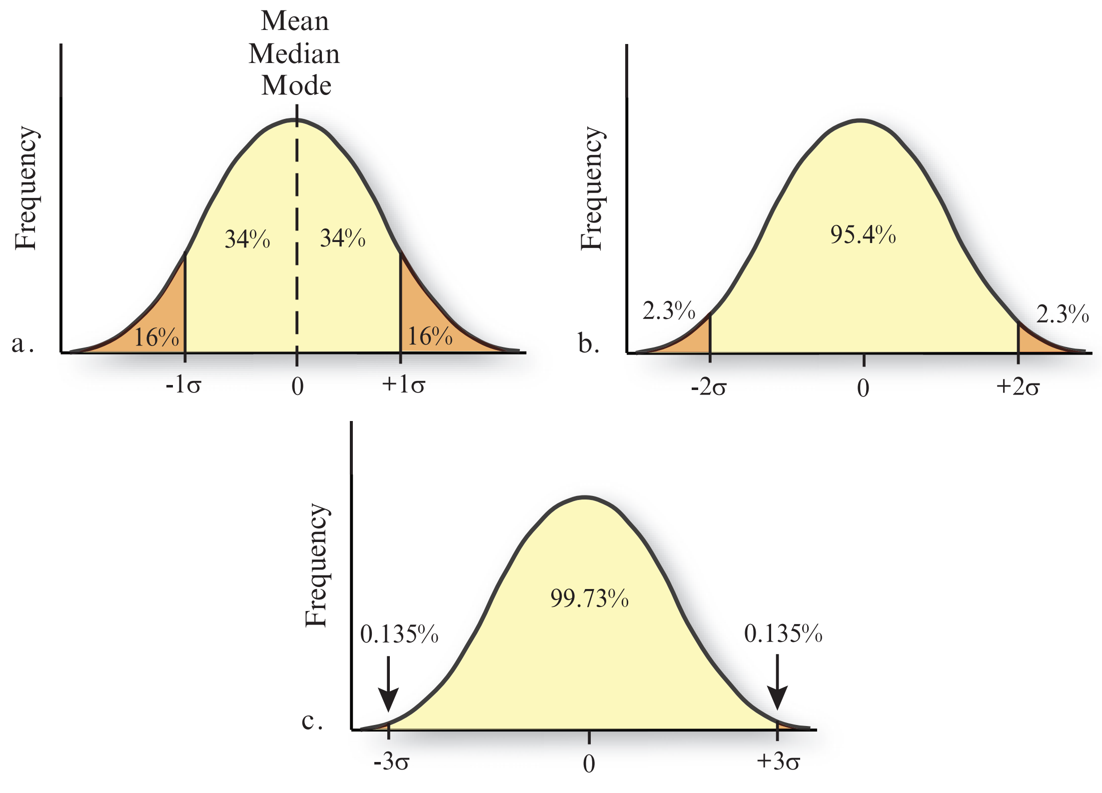


FIGURE 4–6 The area under the normal curve for a) 1, b) 2, and c) 3 standard deviations.

systems with delicate electronics that can experience spikes in voltage and other unfortunate malfunctions. When unusual values are not encountered, the range is a very important statistic often used in image enhancement functions such as min–max contrast stretching discussed in Chapter 8.

The **variance** of a sample is the average squared deviation of all possible observations from the sample mean. The variance of a band of imagery, var_k, is computed using the equation:

$$var_k = \frac{\sum_{i=1}^{n} (BV_{ik} - \mu_k)^2}{n} \tag{4.3}$$

The numerator of the expression, $\sum (BV_{ik} - \mu_k)^2$, is the corrected sum of squares (**SS**) (Davis, 2002). If the sample mean (μ_k) was actually the population mean, this would be an accurate measurement of the variance. Unfortunately, there is some underestimation when variance is computed using Equation 4.3 because the sample mean μ_k in Equation 4.1) was calculated in a manner that minimized the squared deviations about it. Therefore, the denominator of the variance equation is reduced to $n - 1$, producing a somewhat larger, unbiased estimate of the sample variance:

$$var_k = \frac{\mathbf{SS}}{n-1}. \tag{4.4}$$

The **standard deviation** is the positive square root of the variance. The standard deviation of the pixel brightness values in a band of imagery, s_k, is computed as:

$$s_k = \sqrt{var_k} \tag{4.5}$$

A small standard deviation suggests that observations are clustered tightly around a central value. Conversely, a large standard deviation indicates that values are scattered widely about the mean. The total area underneath a distribution curve is equal to 1.00 (or 100%). For normal distributions, 68% of the observations lie within –1 standard deviation of the mean (Figure 4-6a), 95.4% of all observations lie within –2 standard deviations (Figure 4-6b), and 99.73% within –3 standard deviations (Figure 4-6c).

The standard deviation is a statistic commonly used in digital image processing (e.g., linear contrast enhancement, spatial filtering, parallelepiped classification,

TABLE 4–1 A sample dataset of brightness values used to demonstrate the computation of a variance–covariance matrix

Pixel	Band 1 (green)	Band 2 (red)	Band 3 (near-infrared)	Band 4 (near-infrared)
(1,1)	130	57	180	205
(1,2)	165	35	215	255
(1,3)	100	25	135	195
(1,4)	135	50	200	220
(1,5)	145	65	205	235

and accuracy assessment) (e.g., Schowengerdt, 2007; Russ, 2011; Richards, 2013; Chityala and Pudipeddi, 2014). To interpret variance and standard deviation, analysts should not attach a significance to each numerical value but should compare one variance or standard deviation to another. The sample having the largest variance or standard deviation has the greater spread among the brightness values of the observations, provided all the measurements were made in the same units.

Measures of Distribution (Histogram) Asymmetry and Peak Sharpness

Sometimes it is useful to compute additional statistical measures that describe in quantitative terms various characteristics of the distribution (histogram). **Skewness** is a measure of the asymmetry of a histogram and is computed using the formula (Pratt, 2007):

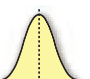
$$skewness_k = \frac{\sum_{i=1}^{n}\left(\frac{BV_{ik}-\mu_k}{s_k}\right)^3}{n} . \qquad (4.6)$$

A perfectly symmetric histogram has a skewness value of zero.

A histogram can be symmetric but have a peak that is very sharp or one that is subdued when compared with a perfectly normal distribution. A perfectly normal frequency distribution (histogram) has zero **kurtosis**. The greater the positive kurtosis value, the sharper the peak in the distribution when compared with a normal histogram. Conversely, a negative kurtosis value suggests that the peak in the histogram is less sharp than that of a normal distribution. Kurtosis is computed using the formula:

$$kurtosis_k = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{BV_{ik}-\mu_k}{s_k}\right)^4\right] - 3 . \qquad (4.7)$$

Outliers or blunders in the remotely sensed data can have a serious impact on the computation of skewness and kurtosis. Therefore, it is desirable to remove (or repair) bad data values before computing skewness or kurtosis.



Multivariate Image Statistics

Remote sensing research is often concerned with the measurement of how much radiant flux is reflected or emitted from an object in more than one band (e.g., in red and near-infrared bands). It is useful to compute multivariate statistical measures such as covariance and correlation among the several bands to determine how the measurements covary. Later it will be shown that variance–covariance and correlation matrices are used in remote sensing principal components analysis (PCA), feature selection, classification and accuracy assessment (e.g., Bishop, 2007; Congalton and Green, 2008; Prost, 2013; Richards, 2013) (Chapters 8 through 13). For this reason, we will examine how the variance–covariance between bands is computed and then proceed to compute the correlation between bands. Although performed on a simple dataset consisting of just five hypothetical pixels (Table 4-1), this example provides insight into the utility of these statistics for digital image processing purposes. Later, these statistics are computed for a seven-band Charleston, SC, Thematic Mapper scene consisting of 240 × 256 pixels. Note that this scene is much larger than the 120 × 100 pixel subscene in Figure 4-2a.

The following examples are based on an analysis of the first five pixels [(1, 1), (1, 2), (1, 3), (1, 4) and (1, 5)] in a four-band (green, red, near-infrared, near-infrared) hypothetical multispectral dataset obtained over vegetated terrain. Thus, each pixel consists of four spectral measurements (Table 4-1). Note the low brightness values in band 2 caused by plant chlorophyll absorption of red light for photosynthesis. Increased reflectance of

TABLE 4–2 Univariate statistics for the hypothetical sample dataset.

Band	1	2	3	4
Mean (μ_k)	135.00	46.40	187.0	222.00
Standard deviation (s_k)	23.71	16.27	31.4	23.87
Variance (var_k)	562.50	264.80	1007.5	570.00
Minimum (min_k)	100.00	25.00	135.0	195.00
Maximum (max_k)	165.00	65.00	215.0	255.00
Range (BV_r)	65.00	40.00	80.0	60.00

the incident near-infrared energy by the green plant results in higher brightness values in the two near-infrared bands (3 and 4). Although it is a small hypothetical sample dataset, it represents well the spectral characteristics of healthy green vegetation.

The simple univariate statistics for such data are usually reported as shown in Table 4-2. In this example, band 2 exhibits the smallest variance (264.8) and standard deviation (16.27), the lowest brightness value (25), the smallest range of brightness values (65 – 25 = 40), and the lowest mean value (46.4). Conversely, band 3 has the largest variance (1007.5) and standard deviation (31.74) and the largest range of brightness values (215 – 135 = 80). These univariate statistics are of value but do not provide useful information concerning whether the spectral measurements in the four bands vary together or are completely independent.

Covariance in Multiple Bands of Remote Sensor Data

The different remote sensing–derived spectral measurements for each pixel often change together in a predictable fashion. If there is no relationship between the brightness value in one band and that of another for a given pixel, the values are mutually independent; that is, an increase or decrease in one band's brightness value is not accompanied by a predictable change in another band's brightness value. Because spectral measurements of individual pixels may not be independent, some measure of their mutual interaction is needed. This measure, called the **covariance**, is the joint variation of two variables about their common mean. To calculate covariance, we first compute the corrected sum of products (**SP**) defined by the equation (Davis, 2002):

TABLE 4–3 Format of a variance–covariance matrix.

	Band 1	Band 2	Band 3	Band 4
Band 1	$\mathbf{SS_1}$	$cov_{1,2}$	$cov_{1,3}$	$cov_{1,4}$
Band 2	$cov_{2,1}$	$\mathbf{SS_2}$	$cov_{2,3}$	$cov_{2,4}$
Band 3	$cov_{3,1}$	$cov_{3,2}$	$\mathbf{SS_3}$	$cov_{3,4}$
Band 4	$cov_{4,1}$	$cov_{4,2}$	$cov_{4,3}$	$\mathbf{SS_4}$

$$\mathbf{SP}_{kl} = \sum_{i=1}^{n} (BV_{ik} - \mu_k)(BV_{il} - \mu_l) \,. \tag{4.8}$$

In this notation, BV_{ik} is the ith measurement of band k, and BV_{il} is the ith measurement of band l with n pixels in the study area. The means of bands k and l are μ_k and μ_l, respectively. In our example, variable k might stand for band 1 and variable l could be band 2. It is computationally more efficient to use the following formula to arrive at the same result:

$$\mathbf{SP}_{kl} = \sum_{i=1}^{n} (BV_{ik} \times BV_{il}) - \frac{\sum_{i=1}^{n} BV_{ik} \sum_{i=1}^{n} BV_{il}}{n} \,. \tag{4.9}$$

The quantity is called the uncorrected sum of products. The relationship of $\mathbf{SP}_{kl}$ to the sum of squares (**SS**) can be seen if we take k and l as being the same, that is:

$$\mathbf{SP}_{kk} = \sum_{i=1}^{n} (BV_{ik} \times BV_{ik}) - \frac{\sum_{i=1}^{n} BV_{ik} \sum_{i=1}^{n} BV_{ik}}{n} \tag{4.10}$$

$$= \mathbf{SS}_k \,.$$

Just as simple variance was calculated by dividing the corrected sums of squares (**SS**) by $(n - 1)$, covariance is calculated by dividing **SP** by $(n - 1)$. Therefore, the covariance between brightness values in bands k and l, cov_{kl}, is equal to (Davis, 2002):

$$\mathbf{SS}_k = \frac{\mathbf{SP}_{kl}}{n-1} \,. \tag{4.11}$$

The sums of products (**SP**) and sums of squares (**SS**) can be computed for all possible combinations of the four spectral variables in Table 4-1. These data can be arranged in a 4 × 4 variance–covariance matrix, as

TABLE 4–4 Variance–covariance matrix of the sample data.

	Band 1	Band 2	Band 3	Band 4
Band 1	562.50			
Band 2	135.00	264.80		
Band 3	718.75	275.25	1007.50	
Band 4	537.50	64.00	663.75	570.00

shown in Table 4-3. All elements in the matrix not on the diagonal have one duplicate (e.g., $\text{cov}_{1,2} = \text{cov}_{2,1}$ so that $\text{cov}_{kl} = \text{cov}_{lk}$).

The computation of variance for the diagonal elements of the matrix and covariance for the off-diagonal elements of the data is shown in Table 4-4. The manual computation of the covariance between band 1 and band 2 is shown in Table 4-5.

Correlation between Multiple Bands of Remotely Sensed Data

To estimate the degree of interrelation between variables in a manner not influenced by measurement units, **Pearson's product-moment correlation coefficient** (r) is commonly computed (Samuels et al., 2011; Konishi, 2014). For example, the correlation between two bands (k and l) of remotely sensed data, r_{kl}, is the ratio of their covariance (cov_{kl}) to the product of their standard deviations ($s_k s_l$):

$$r_{kl} = \frac{cov_{kl}}{s_k s_l}. \tag{4.12}$$

Because the correlation coefficient is a ratio, it is a unitless number. Covariance can equal but cannot exceed the product of the standard deviation of its variables, so correlation ranges from +1 to –1. A correlation coefficient of +1 indicates a positive, perfect relationship between the brightness values in two of the bands (i.e., as one band's pixels increase in value, the other band's values also increase in a systematic fashion). Conversely, a correlation coefficient of –1 indicates that the two bands are perfectly inversely related (i.e., as brightness values in one band increase, corresponding pixels in the other band systematically decrease in value).

A continuum of less-than-perfect relationships exists between correlation coefficients of –1 and +1. A correlation coefficient of zero suggests that there is no linear relationship between the two bands of remote sensor data.

TABLE 4–5 Computation of covariance between bands 1 and 2 of the sample data using Equations 4.9 and 4.11.

Band 1	(Band 1 × Band 2)	Band 2
130	7,410	57
165	5,775	35
100	2,500	25
135	6,750	50
145	9,425	65
675	31,860	232

where $\mathbf{SP}_{1,2} = (31{,}860) - \frac{(675)(232)}{5} = 540$

$\text{cov}_{1,2} = \frac{540}{4} = 135$

If we square the correlation coefficient (r_{kl}), we obtain the sample **coefficient of determination** (r^2), which expresses the proportion of the total variation in the values of "band l" that can be accounted for or explained by a linear relationship with the values of the random variable "band k." Thus, a correlation coefficient (r_{kl}) of 0.70 results in an r^2 value of 0.49, meaning that 49% of the total variation of the values of "band l" in the sample is accounted for by a linear relationship with values of "band k."

The between-band correlations are usually stored in a correlation matrix, such as the one shown in Table 4-6 that contains the between-band correlations of our sample data. Usually, only the correlation coefficients below the diagonal are displayed because the diagonal terms have a value of 1.0 and the terms above the diagonal are duplicates.

TABLE 4–6 Correlation matrix of the sample data.

	Band 1	Band 2	Band 3	Band 4
Band 1	—			
Band 2	0.35	—		
Band 3	0.95	0.53	—	
Band 4	0.94	0.16	0.87	—

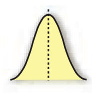
TABLE 4–7 Statistics for the Charleston, SC, Landsat Thematic Mapper scene composed of seven bands of 240 × 256 pixels each.

Band Number (μm)	1 0.45 – 0.52	2 0.52 – 0.60	3 0.63 – 0.69	4 0.76 – 0.90	5 1.55 – 1.75	7 2.08 – 2.35	6 10.4 – 12.5
Univariate Statistics							
Mean	64.80	25.60	23.70	27.30	32.40	15.00	110.60
Standard deviation	10.05	5.84	8.30	15.76	23.85	12.45	4.21
Variance	100.93	34.14	68.83	248.40	568.84	154.92	17.78
Minimum	51.00	17.00	14.00	4.00	0.00	0.00	90.00
Maximum	242.00	115.00	131.00	105.00	193.00	128.00	130.00
Variance–Covariance Matrix							
1	100.93						
2	56.60	34.14					
3	79.43	46.71	68.83				
4	61.49	40.68	69.59	248.40			
5	134.27	85.22	141.04	330.71	568.84		
7	90.13	55.14	86.91	148.50	280.97	154.92	
6	23.72	14.33	22.92	43.62	78.91	42.65	17.78
Correlation Matrix							
1	1.00						
2	0.96	1.00					
3	0.95	0.96	1.00				
4	0.39	0.44	0.53	1.00			
5	0.56	0.61	0.71	0.88	1.00		
7	0.72	0.76	0.84	0.76	0.95	1.00	
6	0.56	0.58	0.66	0.66	0.78	0.81	1.00

In this hypothetical example, brightness values of band 1 are highly correlated with those of bands 3 and 4, that is, $r > 0.94$. A high correlation suggests substantial redundancy in the information content among these bands. Perhaps one or more of these bands could be deleted from the analysis to reduce subsequent computation. Conversely, the relatively lower correlation between band 2 and all other bands suggests that this band provides some type of unique information not found in the other bands. More sophisticated methods

of selecting the most useful bands for analysis are described in later sections.

The results of performing a typical statistical analysis program on the Charleston, SC, Landsat TM data are summarized in Table 4-7. Band 1 exhibits the greatest range of brightness values (from 51 to 242) due to Rayleigh and Mie atmospheric scattering of blue wavelength energy. The near- and middle-infrared bands (4, 5, and 7) all have minimums near or at zero. These values are low because much of the Charleston scene is composed of open water, which absorbs much of the incident near- and middle-infrared radiant flux, thereby causing low reflectance in these bands. Bands 1, 2, and 3 are all highly correlated with one another ($r >$ 0.95), indicating that there is substantial redundant spectral information in these channels. Although not to the same degree, there is also considerable redundancy among the reflective and middle-infrared bands (4, 5, and 7), as they exhibit correlations ranging from 0.66 to 0.95. Not surprisingly, the lowest correlations occur when a visible band is compared with an infrared band, especially bands 1 and 4 ($r = 0.39$). In fact, band 4 is the least redundant infrared band when compared with all three visible bands (1, 2, and 3). For this reason, TM band 4 (0.76 to 0.90 μm) is used as an example throughout much of the text. As expected, the thermal-infrared band 6 data (10.4 to 12.5 μm) are highly correlated with the middle-infrared bands (5 and 7).



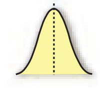
Feature Space Plots

The univariate and multivariate statistics discussed provide accurate, fundamental information about the individual band statistics including how the bands covary and correlate. Sometimes, however, it is useful to examine statistical relationships graphically.

Individual bands of remotely sensed data are often referred to as **features** in the pattern recognition literature. To truly appreciate how two bands (features) in a remote sensing dataset covary and if they are correlated or not, it is often useful to produce a two-band feature space plot.

A two-dimensional **feature space plot** extracts the brightness value for every pixel in the scene in two bands and plots the frequency of occurrence in a 255 by 255 feature space (assuming 8-bit data). The greater the frequency of occurrence of unique pairs of values, the brighter the feature space pixel. For example, we know from Table 4-7 that Landsat Thematic Mapper bands 3 and 4 of the November 9, 1982, Charleston, SC, scene are not highly correlated ($r = 0.53$). A two-dimensional feature space plot of the two bands is shown in Figure 4-7. The original histograms for bands 3 and 4 are also provided. Because the two bands are not highly correlated, the cloud of points looks somewhat like a tilted cap in two-dimensional feature space. The bright areas in the plot represent pixel pairs that have a high frequency of occurrence in the images. It is usually exciting to see feature space plots such as this because it suggests that there is great information of value in these two bands. It reveals visually that the two bands do not contain much redundant information. If the two bands were highly correlated, there would be significant redundant information between the two bands and the cloud of points would appear as a relatively narrow ellipse generally trending diagonally somewhere in the feature space between 0,0 and the 255,255 coordinates (not shown).

The two-dimensional feature space plot also drives home the point that in this particular dataset, the Landsat Thematic Mapper bands 3 and 4 brightness values are seriously constrained to the lower half of the possible 256 values (i.e., band 3 has a maximum value of 131; band 4 has a maximum value of only 4). This suggests that the data in both bands will probably need to be carefully contrast stretched if humans are going to be called upon to visually extract useful information from these individual bands. More will be said about the use of feature space plots in the discussion on feature space partitioning and image classification in Chapter 9.



Geostatistical Analysis, Autocorrelation and Kriging Interpolation

A random variable distributed in space (e.g., spectral reflectance) is said to be regionalized. We can use **geostatistical analysis** to extract the spatial properties of regionalized variables (Curran, 1988; Woodcock et al., 1988ab; Burnicki, 2011). Once quantified, the regionalized variable properties can be used in many remote sensing applications such as image classification (e.g., Maillard, 2003) and the allocation of spatially unbiased sampling sites during classification map accuracy assessment (e.g., Zhu and Stein, 2006; Van der Meer, 2012). An important application of **geostatistics** is the prediction of values at unsampled locations. For example, one could obtain LiDAR-derived elevation values throughout a study area. Although they are very accurate, there may be small or large areas where LiDAR-derived elevation data are missing (i.e., data voids). Geostatistical interpolation techniques could be used to evaluate the spatial relationships associated with the existing LiDAR-derived elevation data to create a new,

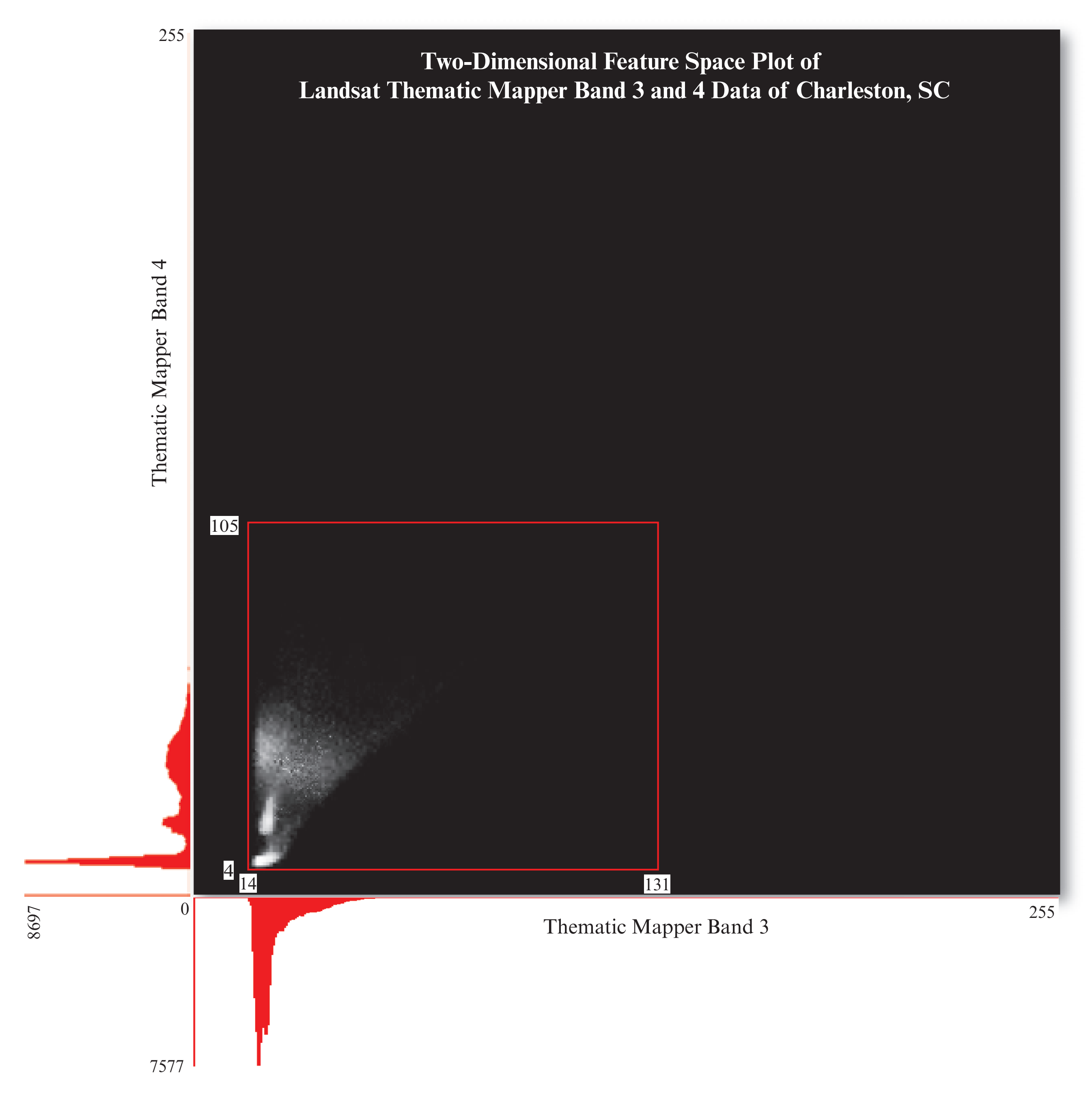


FIGURE 4–7 A two-dimensional feature space plot of Landsat Thematic Mapper bands 3 and 4 data of Charleston, SC, obtained on November 9, 1982. The original histograms of TM bands 3 and 4 are also provided. The minimum and maximum values within each band are bounded by a white box.The greater the frequency of occurrence of a particular pair of values in band 3 and band 4 in the original imagery, the brighter the pixel in the 255 × 255 feature space. For example, it appears that there are many pixels in the scene that have a brightness value of approximately 17 in band 3 and a brightness value of 7 in band 4. Conversely, there is at least one pixel in the scene with a brightness value of 131 in band 3 and 105 in band 4. Because there is probably only one such pair, it shows up as a dark pixel in the feature space.

improved systematic grid of elevation values. This application will be demonstrated.

Geostatistical analysis is a special branch of spatial statistics that takes into account not only the distance between **control point** observations but also their spatial autocorrelation (Kalkhan, 2011; Jensen and Jensen, 2013). Originally, geostatistics was synonymous with kriging—a statistical version of interpolation. **Kriging**, named after the work of Danie Krige (1951), is a generic name for a family of least-squares linear regression algorithms used to estimate the value of a continuous attribute (e.g., elevation) at any unsampled location using only the attribute data available over the

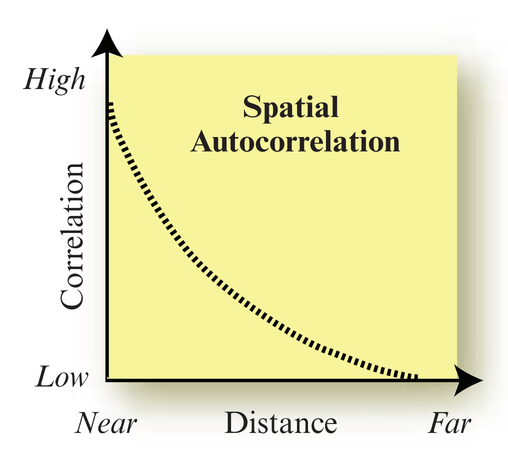


FIGURE 4–8 Phenomena that are geographically closer to one another are generally more highly correlated than phenomena that are farther apart. Geostatistical analysis incorporates spatial autocorrelation information in the Kriging-interpolation process.

study area (Lo and Yeung, 2006; Bachi, 2010). However, geostatistical analysis now includes not only kriging but also the traditional deterministic spatial interpolation methods. One of the essential features of geostatistics is that the phenomenon being studied (e.g., elevation, reflectance, temperature, precipitation, a land cover class) must be capable of existing continuously throughout the landscape.

Most people are aware that things that are close to one another in geographic space are more alike than those that are farther away. Therefore, as distance increases, the spatial autocorrelation decreases as shown in Figure 4-8. Kriging makes use of this spatial autocorrelation information. Kriging is similar to “distance weighted interpolation” in that it *weights* the surrounding nearby values to derive a prediction for each new location in the output raster dataset. However, the weights are based not only on the distance between the measured control points and the point to be predicted (used in Inverse-Distance-Weighting), but also on the overall *spatial arrangement* among the measured points (i.e., their autocorrelation). This is the most significant difference between deterministic (traditional) and geostatistical analysis. Traditional statistical analysis assumes the samples derived for a particular attribute are *independent* and not correlated in any way. Conversely, geostatistical analysis allows users to compute distances between observations and to model autocorrelation as a function of distance and direction. This information is then used to *refine* the kriging interpolation process, making predictions at new locations more accurate than those derived using traditional methods such as Inverse-Distance-Weighting (Lo and Yeung, 2006; Kalkhan, 2011; Esri, 2014).

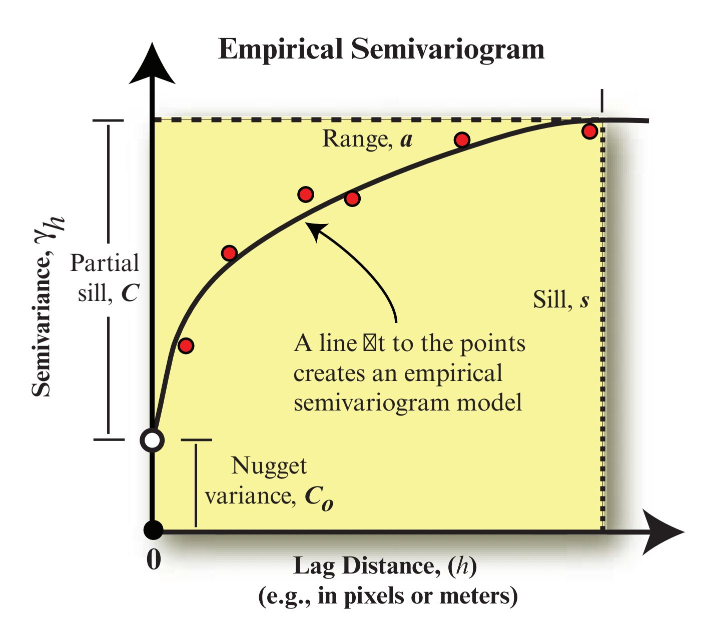


FIGURE 49 The *z*-values of points (e.g., pixels in an image or locations on the ground if collecting *in situ* data) separated by various lag distances (*h*) may be compared and their semivariance (γ_h) computed (based on Isaaks and Srivastava, 1989; Lo and Yeung, 2006; Jensen and Jensen, 2013). The semivariance (γ_h) at each lag distance may be displayed as a *semivariogram* with the range, sill, and nugget variance char-at e is is d id in Tab e 4-9.

The Kriging process generally involves two operations:

- quantifying the spatial structure of the surrounding data points using variography, and
- predicting a $Value_{krig}$ at each new location in the output dataset.

Variography is the process whereby a spatially dependent model is fit to the data and the spatial structure is quantified. To make a prediction for an unknown value at a specific location, Kriging uses the fitted model from variography, the spatial data configuration, and the values of the measured sample points around the prediction location.

One of the most important measurements used to understand the spatial structure of regionalized variables is the **empirical semivariogram** (Figure 4-9), which can be used to relate the semivariance to the amount of spatial separation (and autocorrelation) between samples. The semivariance provides an unbiased description of the scale and pattern of spatial variability throughout a region. For example, if elevation values of a relatively flat area are examined, there may be little spatial variability (variance), which will result in a semivariogram with predictable characteristics. Conversely, the elevation found in rugged terrain may exhibit significant spatial variability resulting in an entirely different semivariogram.

Calculating Average Semivariance

Consider the hypothetical situation where we want to determine the elevation of an unknown point by evaluating the spatial characteristics of six nearby points. The six points, Z_1 through Z_6, are arranged in a line in a Cartesian coordinate system in Table 4-8a. The relationship between a pair of pixels *h* intervals apart (*h* is referred to as the *lag distance*; Table 4-9) can be given by the average variance of the differences between all such pairs. The unbiased estimate of the average semivariance of the population (γ_h) is expressed through the relationship (Isaaks and Srivastava, 1989; Slocum et al., 2008):

$$\gamma_h = \frac{\sum_{i=1}^{n-h}(Z_i - Z_{i+h})^2}{2(n-h)} \tag{4.13}$$

where Z_i is a control point value; *h* is the multiple of the distance between control points, and *n* is the number of points. The total number of possible pairs *m* along the transect is computed by subtracting the lag distance *h* from the total number of pixels present in the dataset *n*, that is, $m = n - h$ (Brivio and Zilioli, 2001; Lo and Yeung, 2006). In this example, we are only computing the semivariance in the *x*-direction. In practice, semivariance is computed for pairs of observations in all directions, N, NE, E, SE, S, SW, W, and NW (Maillard, 2003). In this manner, directional semivariograms are derived and directional influences can be examined.

The computation of the semivariance for the six elevation control points is summarized in Table 4-8b. The semivariance (γ_h) of the six points when $h = 1$ is 17.9 based on the following calculations:

$$\gamma_h = \frac{\sum_{i=1}^{6-1}(Z_i - Z_{i+h})^2}{2(6-1)} \text{ where}$$

$$\gamma_h = (Z_1 - Z_2)^2 + (Z_2 - Z_3)^2 + (Z_3 - Z_4)^2 + (Z_4 - Z_5)^2 + (Z_5 - Z_6)^2 / 10$$

$$\gamma_h = (10 - 15)^2 + (15 - 20)^2 + (20 - 30)^2 + (30 - 35)^2 + (35 - 33)^2 / 10$$

$$\gamma_h = 17.9$$

The average semivariance is a good measure of the amount of *dissimilarity* between spatially separate control points. Generally, the larger the average semivariance (γ_h), the less similar are the point observations. In Table 4-8b we see that at a lag distance of 1, semivariance = 17.9. At a lag distance of 2, semivariance = 69.88. At a lag distance of 3, semivariance =161.5, etc. In this hypothetical dataset, as the lag distance increases, there is greater dissimilarity among the six ground control points being examined. The greater the lag distance, the less correlated the observations.

Empirical Semivariogram

The semivariogram is a plot of the average semivariance value on the *y*-axis with the various lags (*h*) investigated on the *x*-axis, as shown in Figure 4-9 and Table 4-8c. Important characteristics of the semivariogram include:

- lag distance (*h*) on the *x*-axis,
- sill (*s*),
- range (*a*),
- nugget variance (C_o), and
- spatially dependent structure displayed graphically (e.g., Figure 4-9; Table 4-8c).

When spatial correlation exists, pairs of points that are close together (on the far left of the *x*-axis) should have less difference (be low on the *y*-axis). As points become farther away from each other (moving right on the *x*-axis), in general, the difference squared should be greater (moving up on the *y*-axis). The semivariogram model often flattens out at a certain lag distance from the origin. The distance where the model *first* flattens out is the *range*. The range is the distance over which the samples are spatially correlated. The *sill* is the value on the *y*-axis where the range is located. It is the point of maximum variance and is the sum of the structural spatial variance and the nugget effect. The *partial sill* is the sill minus the nugget.

The semivariogram value on the *y*-axis should theoretically equal zero when lag (*h*) equals zero. However, at an infinitesimally small separation distance, the semivariogram often exhibits a *nugget effect*, which is >0. The nugget effect is attributed to measurement errors or spatial sources of variation at distances smaller than the sampling interval (or both).

A semivariogram provides information on the spatial autocorrelation of control point datasets. For this reason and to ensure that Kriging predictions have positive Kriging variances, a model is fit to the semivariogram (i.e., a continuous function or curve). This model quantifies the spatial autocorrelation in the data (Johnston et al., 2001).

TABLE 4–8 Semivariance computation (adapted from Jensen and Jensen, 2013).

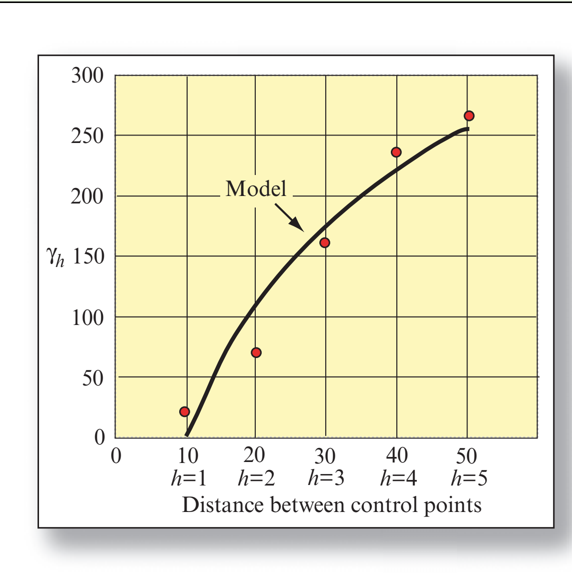
a. Six equally-spaced hypothetical elevation control points (Z_1 through Z_6).

lag =5 — 1 pair
lag =4 — 2 pairs
lag =3 — 3 pairs
lag =2 — 4 pairs
lag =1 — 5 pairs
y
Z_1 Z_2 Z_3 Z_4 Z_5 Z_6
10 15 20 30 35 33
0 10 20
0 10 20 30 40 50 60 x

b. Semivariance computations.

	h				
	1	2	3	4	5
$(Z_1 - Z_{1+h})^2$	25	100	400	625	529
$(Z_2 - Z_{2+h})^2$	25	225	400	324	
$(Z_3 - Z_{3+h})^2$	100	225	169		
$(Z_4 - Z_{4+h})^2$	25	9			
$(Z_5 - Z_{5+h})^2$	4				
$\sum_{i=1}^{n-h} (Z_i - Z_{i+h})^2$	179	559	969	949	529
$2(n-h)$	10	8	6	4	2
γ_h	**17.9**	**69.88**	**161.5**	**237.25**	**264.5**

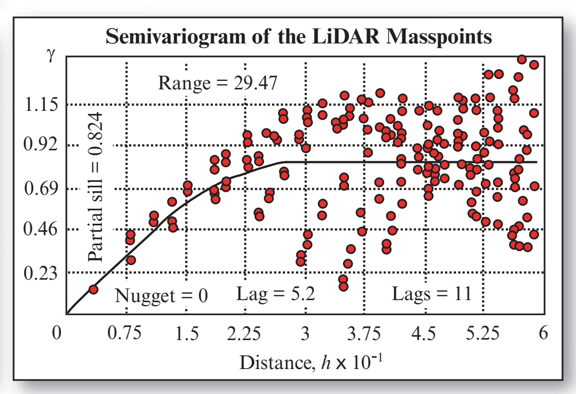
c. Empirical semivariogram model.



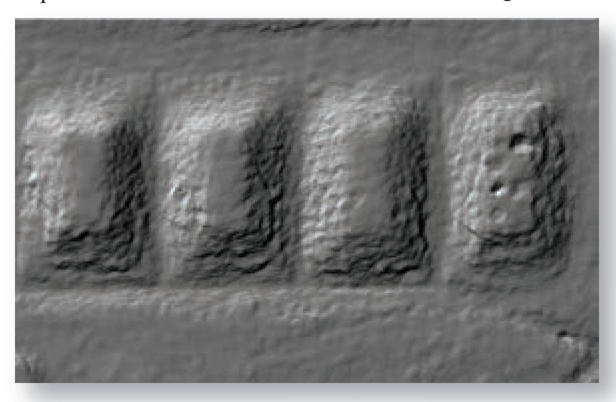
Kriging Spatial Interpolation of LiDAR Masspoints, Analytical Hill-shading, Hypsometric Tinting, and Contour Extraction



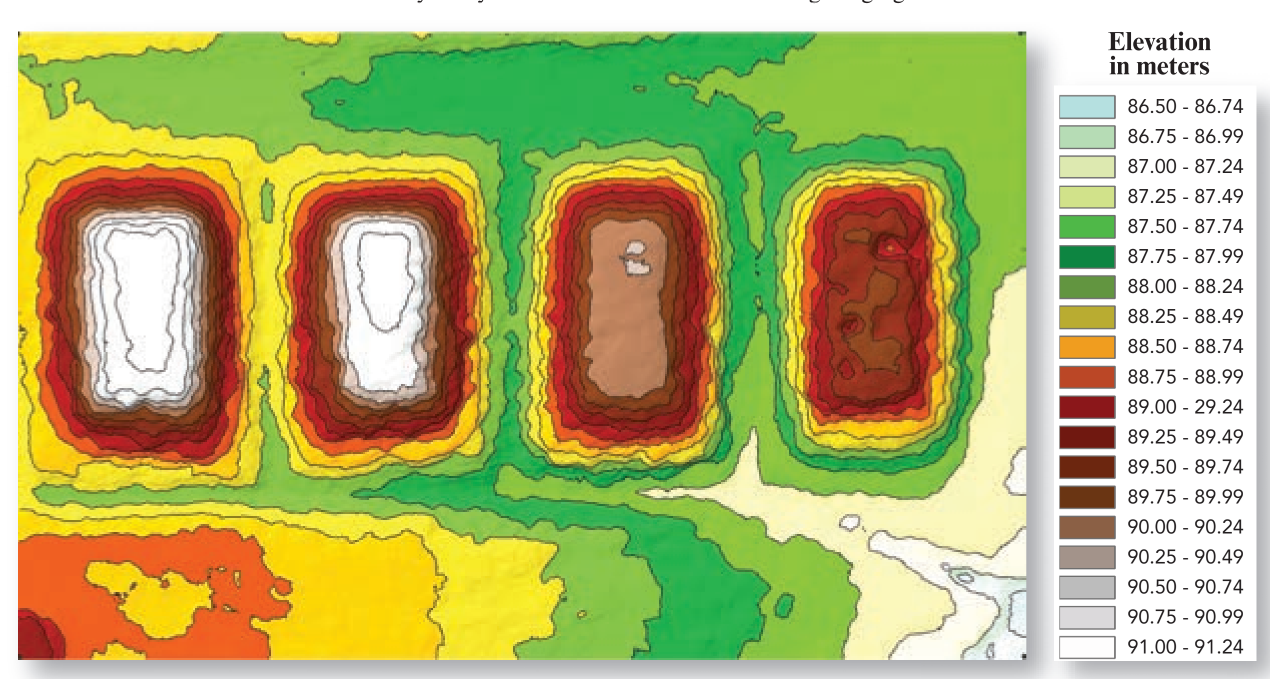
a. LiDAR masspoints.



b. Semivariogram of the LiDAR masspoints.



c. Analytically hill-shaded surface created using Kriging.



d. Kriging-derived elevation surface hypsometrically tinted using 19 class intervals and overlaid with 25 cm contours.

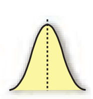
FIGURE 4–10 a) 37,150 LiDAR masspoints (green dots) overlaid on a 1 × 1 m color-infrared orthophotograph. b) A semivariogram of the 37,150 LiDAR masspoints processed using spherical Kriging. This semivariogram contains autocorrelation information derived from all orientations (e.g., N, E, S, and W). c) An analytically hill-shaded surface created using a Kriging model. The cell size is 0.25 × 0.25 m. d) The Kriged surface color-coded using 19 class intervals and overlaid with 25 cm contours derived from the elevation surface.

TABLE 4–9 Terminology and symbols in a typical empirical semivariogram (based on Curran, 1988; Johnston et al., 2001; Lo and Yeung, 2006; Jensen and Jensen, 2013).

Term	Symbol	Definition
Lag	h	The linear (horizontal) distance that separates any two locations (i.e., a sampling pair). A lag has length (distance) and direction (orientation).
Sill	s	The maximum level of the modeled semivariogram. The value that the variogram tends to when lag distances become very large. At large lag distances, variables become uncorrelated, so the sill of the semivariogram is equal to the variance of the random variable.
Range	a	Point on the h-axis where the modeled semivariogram nears a maximum. The distance beyond which there is little or no autocorrelation among variables. Places closer than the range are autocorrelated, places farther apart are not.
Nugget variance	C_o	The location where the modeled semivariogram intercepts the $\gamma(h)$ axis. Represents the independent error, measurement error, or microscale variation at spatial scales that are too fine to detect. The nugget effect is a discontinuity at the origin of the semivariogram model.
Partial sill	C	The sill minus the nugget variance describes the spatially dependent structural variance.

The line fitted to the data points in Figure 4-9 and in Table 4-8c is called a *semivariogram model*. The model is a line through the average semivariance versus average distance between point pairs. The user selects the functional form that best represents the distribution (e.g., spherical, circular, etc.). The coefficients of the function are empirically derived from the data.

It is possible to work with relatively few or many observations to predict the value at unknown locations using Kriging. For example, consider a relatively large LiDAR dataset consisting of 37,150 masspoints located on the Savannah River Site shown in Figure 4-10a. The semivariogram shown in Figure 4-10b contains detailed information about the range, partial sill, and nugget associated with the entire dataset. A spherical geostatistical Kriging algorithm was used to create the raster continuous surface (Figure 4-10c) from the 37,150 LiDAR masspoints shown in Figure 4-10a. The raster surface has been analytically hill-shaded. A color-coded version of the Kriged elevation surface with 19 class intervals is shown in Figure 4-10d overlaid with derived 25 cm contours.



References

Bachi, R., 2010, *New Methods of Geostatistical Analysis and Graphical Presentation: Distributions of Populations over Territories*, NY: Kluwer Academic, 478 p.

Bishop, C. M., 2007, *Pattern Recognition and Machine Learning*, NY: Springer, 738 p.

Bossler, J. D. (Ed.), 2010, *Manual of Geospatial Science and Technology*, 2nd Ed., Boca Raton: CRC Press, 854 p.

Brivio, P. A. and E. Zilioli, 2001, "Urban Pattern Characterization Through Geostatistical Analysis of Satellite Images," in Donnay, J. P., Barnsley, M. J., and P. A. Longley (Eds.), *Remote Sensing and Urban Analysis*, London: Taylor & Francis, 39–53.

Burnicki, A. C., 2011, "Spatio-temporal Errors in Land-Cover Change Analysis: Implications for Accuracy Assessment," *International Journal of Remote Sensing*, 32(22):7487–7512.

Chityala, R. and S. Pudipeddi, 2014, *Image Processing and Acquisition using Python*, Boca Raton: CRC Press, 390 p.

Congalton, R. G. and K. Green, 2008, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices,* 2nd Ed., Boca Raton: Lewis, 183 p.

Curran, P. J., 1988, "The Semivariogram in Remote Sensing: an Introduction," *Remote Sensing of Environment*, 24:493–507.

Davis, J. C., 2002, *Statistics and Data Analysis in Geology*, 3rd ed., NY: John Wiley & Sons, 638 p.

Esri, 2014, *ArcGIS Geostatistical Analyst*, Redlands: Esri, Inc. (www.Esri.com/software/arcgis/extensions/geostatistical/index.html).

FGDC, 2014, *The National Spatial Data Infrastructure*, Washington: Federal Geographic Data Committee, https://www.fgdc.gov/nsdi/nsdi.html.

Gonzalez, R. C. and R. E. Woods, 2007, *Digital Image Processing*, 3rd Ed., NY: Addison-Wesley, 797 p.

Isaaks, E. H. and R. M. Srivastava, 1989, *An Introduction to Applied Geostatistics*, Oxford: The Oxford University Press, 561 p.

Jensen, J. R. and R. R. Jensen, 2013, *Introductory Geographic Information Systems*, Boston: Pearson, 400 p.

Jensen, R. R. and J. M. Shumway, 2010, "Sampling Our World," in Gomez, B., and J. P. Jones III (Eds.), *Research Methods in Geography*, NY: Wiley Blackwell, 77–90.

Johnston, K., Ver Hoef, J. M., Krivoruchko, K., and N. Lucas, 2001, *Using ArcGIS Geostatistical Analyst*, Redlands: Environmental Sciences Research Institute, 300 p.

Kalkhan, M. A., 2011, *Spatial Statistics: GeoSpatial Information Modeling and Thematic Mapping*, Boca Raton: CRC Press, 166 p.

Konishi, S., 2014, *Introduction to Multivariate Analysis: Linear & Nonlinear Modeling*, Boca Raton: CRC Press, 338 p.

Krige, D. G., 1951, *A Statistical Approach to Some Mine Valuations and Allied Problems at the Witwatersrand*, Master's Thesis, University of Witwatersrand, South Africa.

Lanter, D. P., 1991, "Design of a Lineage-based Meta-database for GIS," *Cartography and Geographic Information Systems*, 18(4):255–261.

Lo, C. P. and A. K. W. Yeung, 2006, *Concepts and Techniques of Geographic Information Systems*, 2nd Ed., Boston: Pearson, 493 p.

Maillard, P., 2003, "Comparing Texture Analysis Methods through Classification," *Photogrammetric Engineering & Remote Sensing*, 69(4):357–367.

Papp, J., 2010, *Quality Management in the Imaging Sciences*, 4th Ed., NY: Mosby Elsevier, 352 p.

Pratt, W. K., 2007, *Digital Image Processing*, 4th Ed., NY: John Wiley & Sons, 782 p.

Prost, G. L., 2013, *Remote Sensing for Geoscientists: Image Analysis and Integration*, 3rd Ed., Boca Raton: CRC Press, 702 p.

Richards, J. A. 2013, *Remote Sensing Digital Image Analysis: An Introduction*, 5th Ed., NY: Springer-Verlag, 494 p.

Russ, J. C., 2011, *The Image Processing Handbook*, 6th Ed., Boca Raton: CRC Press, 867 p.

Samuels, M. L. Witmer, J. A. and A. Schaffner, 2011, *Statistics for the Life Sciences*, 4th Ed., Boston: Pearson, 672 p.

Schowengerdt, R. A., 2007, *Remote Sensing: Models and Methods for Image Processing*, 3rd Ed., San Diego: Academic Press, 515 p.

Slocum, T. A., McMaster, R. B., Kessler, F. C. and H. H. Howard, 2008, *Thematic Cartography and Geographic Visualization*, 3rd Ed., Boston: Pearson, 520 p.

Van der Meer, 2012, "Remote-sensing Image Analysis and Geostatistics," *International Journal of Remote Sensing,* 33(18):5644-5676.

Woodcock, C. E., Strahler, A. H. and D. L. B. Jupp, 1988a, "The Use of Variograms in Remote Sensing: I. Scene Models and Simulated Images," *Remote Sensing of Environment*, 25:323–348.

Woodcock, C. E., Strahler, A. H. and D. L. B. Jupp, 1988b, "The Use of Variograms in Remote Sensing: II. Real Images," *Remote Sensing of Environment*, 25:349–379.

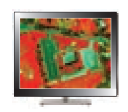
Zhu, Z. Y. and M. L. Stein, 2006, "Spatial Sampling Design for Prediction with Estimated Parameters," *Journal of Agricultural, Biological, and Environmental Statistics*, 11:24–44.

5 DISPLAY ALTERNATIVES AND SCIENTIFIC VISUALIZATION



Scientists interested in displaying and analyzing remotely sensed data actively participate in **scientific visualization**, defined as "visually exploring data and information in such a way as to gain understanding and insight into the data." The difference between scientific visualization and presentation graphics is that the latter are primarily concerned with the communication of information and results that are already understood. During scientific visualization we are seeking to *understand* the data and gain new insight (Earnshaw and Wiseman, 1992; Slocum et al., 2008; Myler, 2013).

Scientific visualization of remotely sensed data is maturing. Its origin can be traced to the simple plotting of points and lines and contour mapping (Figure 5-1). We now have the ability to conceptualize and visualize remotely sensed images in two-dimensional space in true color as shown (two-dimensional to two-dimensional). It is also possible to drape remotely sensed data over a digital elevation model (DEM) and display the synthetic three-dimensional model on a two-dimensional map or computer screen (i.e., three-dimensional to two-dimensional). If we transformed this same three-dimensional model into a physical model that we could touch, it would occupy the three-dimensional to three-dimensional portion of scientific visualization mapping space.



Overview

This chapter identifies the challenges and limitations associated with displaying remotely sensed data and makes suggestions about how to display and visualize the data using black-and-white and color output devices. The characteristics of 8- and 24-bit color look-up tables are presented. Methods of selecting the most appropriate bands to display in a color composite display are presented (e.g., optimum index factor, Sheffield Index). Methods of merging (fusing) different types of remote sensor data are discussed (e.g., component substitution) along with several color coordinate systems (e.g., RGB, intensity–hue–saturation [IHS], and chromaticity). Methods of extracting distance and polygon perimeter, area, and shape information from remotely sensed images are discussed.

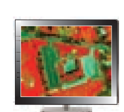


Image Display Considerations

Humans are very adept at visually interpreting continuous-tone images every day as they read magazines and newspapers or watch television (Ready, 2013). Our goal is to capitalize on this talent by providing remotely sensed data in a format that can be easily visualized and interpreted to gain new insight about the Earth (Jensen and Jensen, 2013). The first problem is that the remotely sensed data collected by government agencies (e.g., NASA Landsat 8 data) or private industry (e.g., GeoEye, Inc., DigitalGlobe, Inc., and SPOT Image, Inc.) are in a digital format. How do we convert the brightness values (BVs) stored on any type of storage device into an image that begins to approximate the continuous-tone photographs so familiar to humans? The answer is the creation of a brightness map, also commonly referred to as a grayscale or color image.

A **brightness map** is a computer graphic display of the brightness values, $BV_{i,j,k}$, found in digital remotely

Scientific Visualization Mapping Space

Dimensionality of the Visual Representation: 0D, 1D, 2D, 3D

Dimensionality of the Conceptual Domain: 1D, 2D, 3D, nD

0D: Points | Scatterplots, Particles, Dotsurfaces | Scatterplots, Particle tracers, Dot surfaces

1D: Lines, Curves | Contour maps, Field vectors, Space curves | 3D vector nets, Field vectors, Ribbons

2D: Images: - pseudo-color - true color, Height fields, Contour maps | 3D digital terrain, Tiled surfaces, Stacked textures, Ribbons | Attribute mapping

3D: Volume rendering, Solids modeling | Icons, Attribute mapping

FIGURE 5–1 Scientific visualization mapping space. The *x*-axis is the conceptual domain or how we conceive the information in our minds. The *y*-axis is the actual number of dimensions used to visually represent our conceptual ideas (based on Earnshaw and Wiseman, 1992).

sensed data (refer to Figure 2-1). Ideally, there is a one-to-one relationship between input brightness values and the resultant intensities of the output brightness values on the display as shown in Figure 5-2a. For example, an input *BV* of 0 would result in a very dark (black) intensity on the output brightness map, while a *BV* of 255 would produce a bright (white) intensity. All brightness values between 0 and 255 would be displayed as a continuum of grays from black to white. In such a system, an input brightness value of 127 would be displayed exactly as 127 (mid-gray) in the output image, as shown in Figure 5-2a (assuming contrast stretching does not take place). Unfortunately, it is not always easy to maintain this ideal relationship. In the past, it was common for analysts to have access to devices that displayed only a relatively small range of brightness values (e.g., <50) (Figure 5-2b). In this example, several input brightness values around *BV* 127 might be assigned the same output brightness value of 25. When this occurs, the original remotely sensed data are generalized when displayed, and valuable information may never be seen by the image analyst. Therefore, it is important that whenever possible the one-to-one relationship between input and output brightness values be maintained.

This chapter describes the creation of remote sensing brightness maps using two fundamentally different output devices: hard-copy displays and temporary video displays. **Hard-copy displays** are based on the use of line printers, line plotters, ink-jet printers, laser printers, or film writers to produce tangible hard copies of the imagery for visual examination. **Temporary video displays** are based on the use of black-and-white or color video technology, which displays a temporary image for visual examination. The temporary image can be discarded or subsequently routed to a hard-copy device if desired.



Black-and-White Hard-Copy Image Display

Hard-copy image displays may be produced using line printer/plotters, laser printers, or ink-jet printers.

Line Printer/Plotter Brightness Maps

At one time, the 6- or 8-line per inch alphanumeric line printer was the device most often used to produce

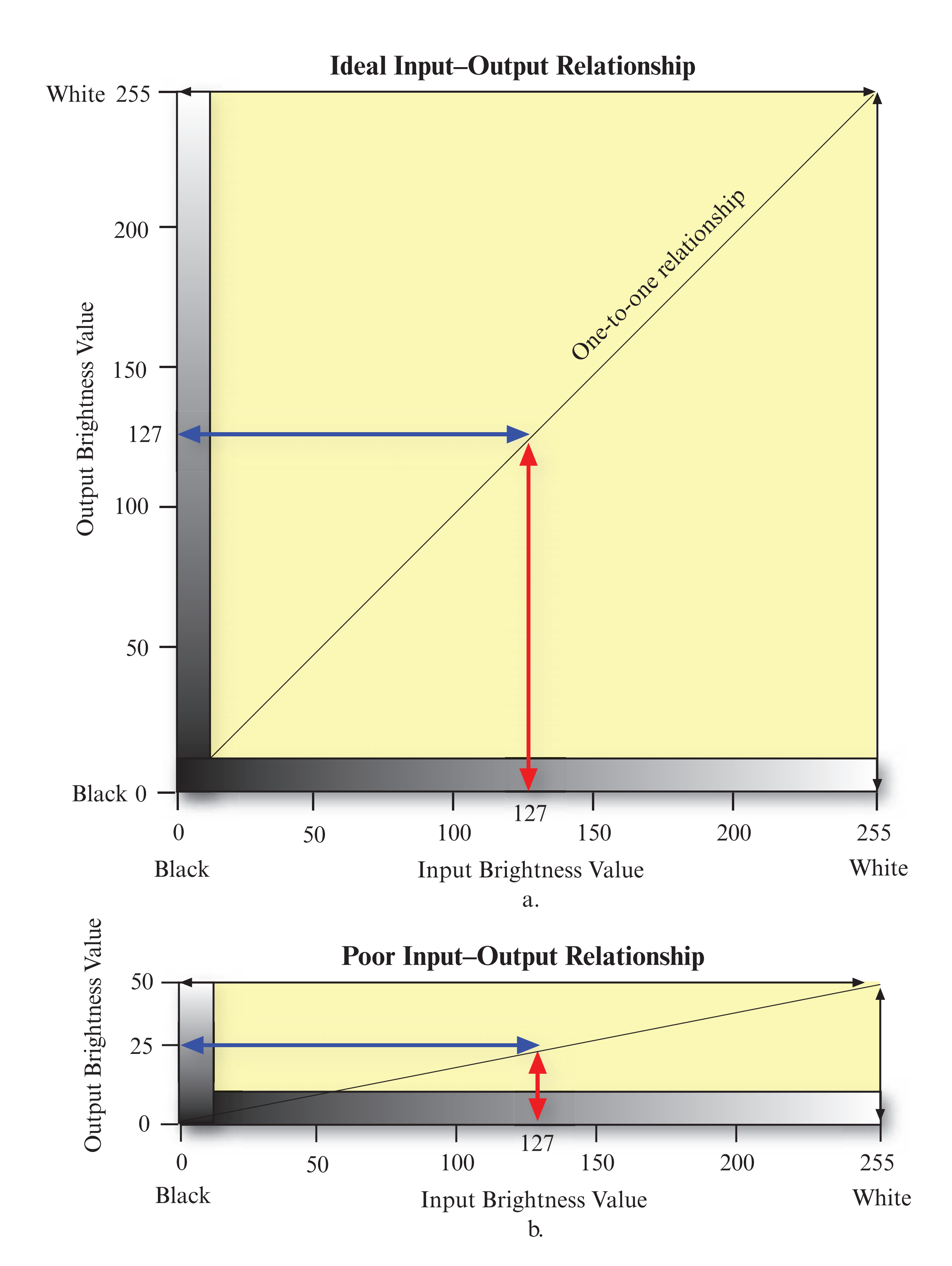


FIGURE 5–2 a) An ideal one-to-one relationship between 8-bit input remote sensing brightness values and the output brightness map. Most modern computer graphic and digital image processing workstation environments maintain this relationship. b) A poor situation where an analyst's output device is not capable of retaining the one-to-one relationship and must generalize the original 8-bit data down to a more manageable number of brightness map classes. In this case, the output device has only 50 classes. If the 8-bit data were uniformly distributed on the *x*-axis, >5× reduction of information would take place using this hypothetical output device. These two examples assume that the remotely sensed data are not contrast stretched.

FIGURE 5–3 The brightness value of a picture element (pixel) is read from mass storage by the central processing unit (CPU). The digital value of the stored pixel is in its proper *i,j* location in the image processor's random-access memory (RAM), often referred to as a video *frame buffer*. The brightness value is then passed through a black-and-white or color look-up table where modifications can be made. The output from the digital color look-up table is passed to a digital-to-analog converter (DAC). The output from the DAC determines the intensity of the signal for the three guns (red, green, and blue) in the back of the monitor that stimulate the phosphors on a computer cathode-ray tube (CRT) or at a specific *x*, *y* location or the transistors in a liquid crystal display (LCD). A DAC is *not* required if a *digital* monitor is used.

scaled hard-copy images. The input data were *density sliced* into a series of discrete class intervals in which each interval corresponded to a specific brightness value range. To produce density-sliced maps on the line printer, it was necessary to select 1) an appropriate number of class intervals, 2) the size or dimension of each class interval, and 3) the alphanumeric symbolization to be assigned to each class interval. Sometimes, more sophisticated line plotters were available. Line plotters were programmed to give the impression of continuous tone using crossed-line shading produced by intertwining two perpendicular sets of equally spaced parallel lines. Alphanumeric printers and line plotters are still used to produce hard-copy output if laser or ink-jet printers are not available.

Laser or Ink-Jet Printer Brightness Maps

Most image analysts now use relatively inexpensive laser or ink-jet printers to output what appears to be continuous-tone black-and-white or color imagery. This is accomplished by the system's software, which develops a functional relationship between a pixel's input brightness value and the amount of laser toner or ink-jet ink that is applied (output) at the appropriate location on the printed page. These relatively inexpensive printers can normally apply the toner or ink-jet ink at 100 to 1,200 dots per inch (dpi).



Temporary Video Image Display

The most effective display of remote sensor data is based on the use of temporary video displays that have improved black-and-white and color brightness map display capabilities.

TABLE 5–1 Compression scheme characteristics for selected raster image file formats (based on Myler, 2013).

Format	Compression Scheme
Raw	None—pixel values are stored directly. For example, ERDAS and ENVI software stores the remote sensor data in its raw, original format with no compression.
GIF (Graphic interchange format)	*Lossless* for grayscale images. Can be *lossy* for color images.
JPEG (Joint photographic experts group)	Inherently *lossy*, JPEG removes redundant data based on a model of human color perception. JPEG can also be used to store a *lossless* image, but the file size will be greater than that of the raw image because additional information is added to the file.
TIFF (Tagged interchange file format)	Compression scheme selectable but most schemes used are *lossless*. TIFF file is usually larger than the raw file because additional information is added to the file.

Black-and-White and Color Brightness Maps

A **video image display** of remotely sensed data can be easily modified or discarded. The central processing unit (CPU) reads the digital remote sensor data from a mass storage device (e.g., hard disk or optical disk) and transfers these data to the image processor's random access memory (RAM) frame buffer (Figure 5-3). The **image processor frame buffer** is a collection of display memory composed of *i* lines by *j* columns and *b* bits that can be accessed sequentially, line by line. Each line of digital values stored in the image processor display memory is continuously scanned by a *read mechanism*. The content of the image processor memory is read every 1/60 of a second, referred to as the *refresh rate* of the system. The brightness values encountered during this scanning process are passed to the color **look-up table** (LUT). An analyst can change the contents of the color look-up table to modify how an individual pixel eventually appears on the computer screen. The contents of the look-up table are then passed to a digital-to-analog converter (DAC) that prepares an analog video signal suitable for display on the cathode-ray tube (CRT) or liquid crystal display (LCD) (Myler, 2013). Thus, the analyst viewing a computer screen is actually looking at the video expression of the digital values stored in the image processor's memory. The refresh rate (1/60 second) is so fast that the analyst does not see any significant amount of flicker on the screen. If a digital display is used (as opposed to an analog display), there is no need for a digital-to-analog converter (DAC) because the digital data are passed directly to the digital display.

Image Data Format and Compression Scheme

The most common bitmapped graphics formats are raw, GIF (Graphic interchange format), JPEG (Joint photographic experts group), and TIFF (Tagged interchange file format). Their characteristics are summarized in Table 5-1 (Myler, 2013). When remote sensor data are stored directly with no compression it is called a **raw image file** and is considered to be *lossless*. Conversely, some image formats use a compression scheme to reduce the amount of data that must be stored. Compression schemes take advantage of the repetition or redundancy in neighboring pixel values. Various types of **image data compression** are presented in Table 5-1. They can be *lossless* or *lossy*. Lossy image data compression schemes almost always provide good data reduction but at the risk of losing valuable spectral information in the image (SPIE, 2013). Therefore, digital image processing analysts are strongly encouraged to use only *lossless* image data compression schemes when storing their remote sensor data (e.g., in raw [ERDAS, ENVI] or uncompressed TIFF format) so they don't lose valuable spectral information.

But how does the entire range of brightness values in a lossy or lossless digital image dataset get displayed properly? This is a function of 1) the number of bits per pixel associated with the bitmapped graphic under consideration, 2) the color coordinate system being used, and 3) the video look-up tables associated with the image processor.

Bitmapped Graphics

The digital image processing industry refers to all raster images that have a pixel brightness value at each row and column in a matrix as being **bitmapped** images. The tone or color of the pixel in the image is a function of the value of the bits or bytes associated with the pixel and the manipulation that takes place in a color look-up table. For example, the simplest bitmapped image is a binary image consisting of just ones (1) and zeros (0). It is useful to describe how a simple binary bitmapped image is encoded and displayed, before discussing how grayscale and color images are encoded and displayed.

The simple scene in Figure 5-4 consists of a house, a very large tree, and two road segments. The geographic area has been digitized into a simple raster of 1s and 0s

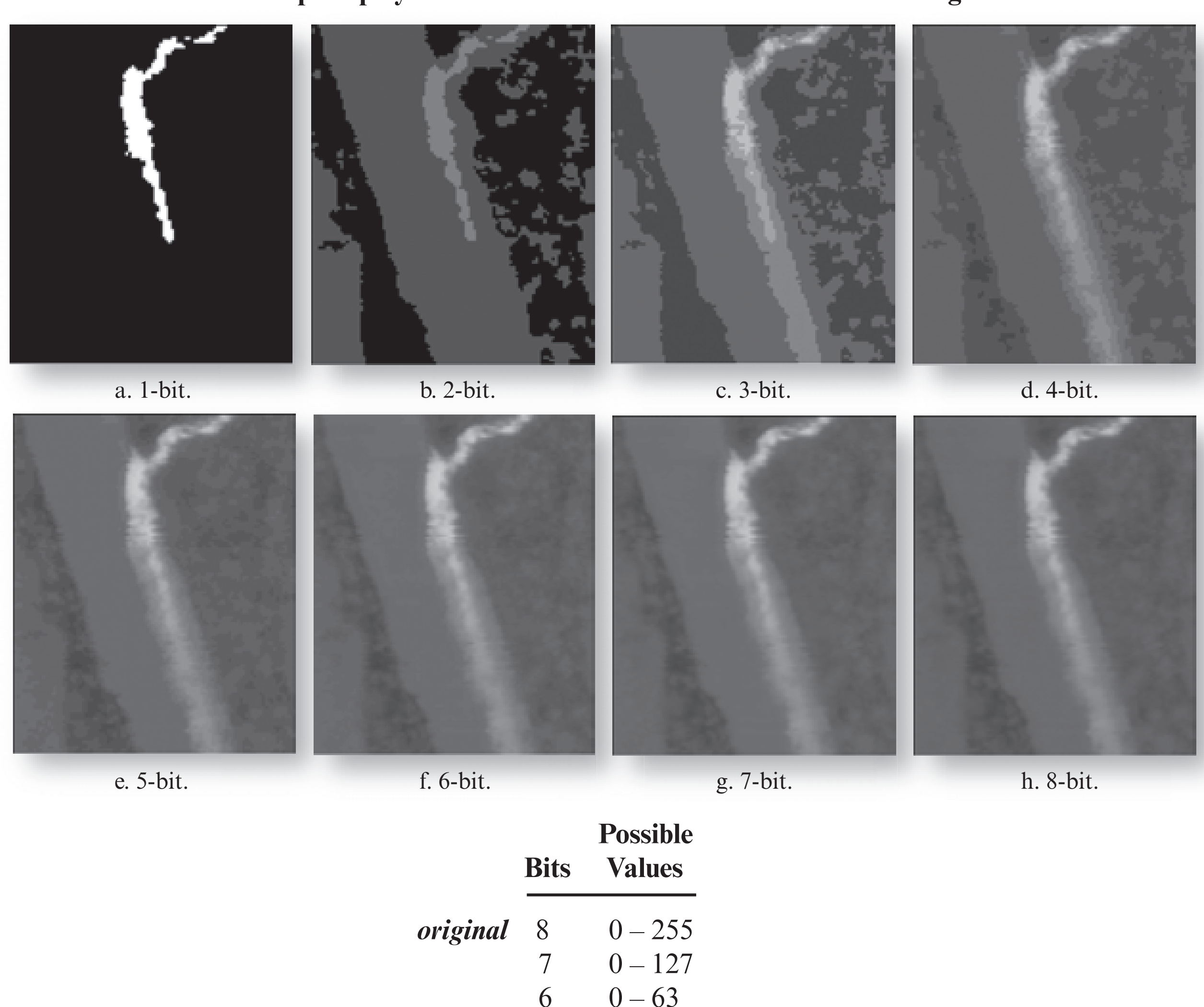
1	0	0	1	1	1	1	1	1	1	1
1	0	0	1	1	1	1	1	1	1	1
1	1	1	1	0	0	0	1	1	1	1
1	1	1	0	0	0	0	0	1	1	1
1	1	0	0	0	0	0	0	0	1	1
1	1	0	0	0	0	0	0	0	1	1
1	1	1	0	0	0	0	0	1	1	0
1	1	1	1	0	0	0	1	1	0	1
1	1	1	1	1	1	1	1	0	1	1
0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1

5. The bytes in the video memory that make up the black-and-white image consist of some bits set to 1 and the rest to 0. Pixels with a 1 are turned on and pixels with a 0 are turned off. The scene consists of 121 pixels, which in a binary black-and-white image can be stored as 16 bytes. The last 7 bits are discarded because they fall outside the matrix of pixels specified in the header.

FIGURE 5–4 Characteristics of a binary (0 and 1), black-and-white *bitmapped* image.

and stored in a bitmapped format as a unique graphics file on a hard disk or other mass storage device. How do we read the contents of the bitmapped graphic file and construct a digital image from the bits and bytes stored in it? First, all raster graphic files contain a header. The first record in the header file is the *signature,* which lets the digital image processing program (application) know what type of file is present. The next record in the header identifies the number of rows (height) and columns (width) in the bitmapped dataset. The final records in the header contain a summary of the number of colors in the palette (e.g., 2) and the nature of the colors in the original palette (e.g., black and white). In this example, the bitmapped image contains 11 rows and 11 columns of data with a black-and-white color palette.

Based on the information in the header, the digital image processing program can extract the values of the 121 individual picture elements in the image. Note that



	Bits	Possible Values
original	8	0 – 255
	7	0 – 127
	6	0 – 63
	5	0 – 31
	4	0 – 15
	3	0 – 7
	2	0 – 3
	1	0 – 1

FIGURE 5–5 Display of the Savannah River thermal infrared data at various bitmap resolutions.

the information content of the entire 121-element matrix can be stored in just 16 bytes (Figure 5-4). Each 8-bit byte contains 8 individual values for 8 pixels. The last 7 bits of byte 16 are not used since the header record specified that there would only be 121 useful values, not the 128 possible values (8 bits × 16 bytes = 128) in a 16-byte dataset. With all the information in hand, the program displays a binary black-and-white image of the scene with zeros in black and ones in white.

Few remotely sensed images of significant value contain only binary (1-bit) information. Instead, most remote sensor data are quantized to 8 bits per pixel with the values ranging from 0 to 255 in each band. For example, consider the various bitmapped displays of the thermal plume in the Savannah River shown in Figure 5-5. As expected, the original 8-bit display contains all of the information in the original dataset with a minimum value of 38 and a maximum value of 188. The most dramatic drop-off in visual information content seems to occur when the data are displayed using 4 bits or fewer. A 4-bit display only records values that have been transformed to be within the range 0 to 15. The binary 1-bit display contains very little useful information. In effect, every pixel in the 1-bit display with a brightness value <128 has a value of 0 and every pixel ≥128 has a value of 1. It is clear from this illustration

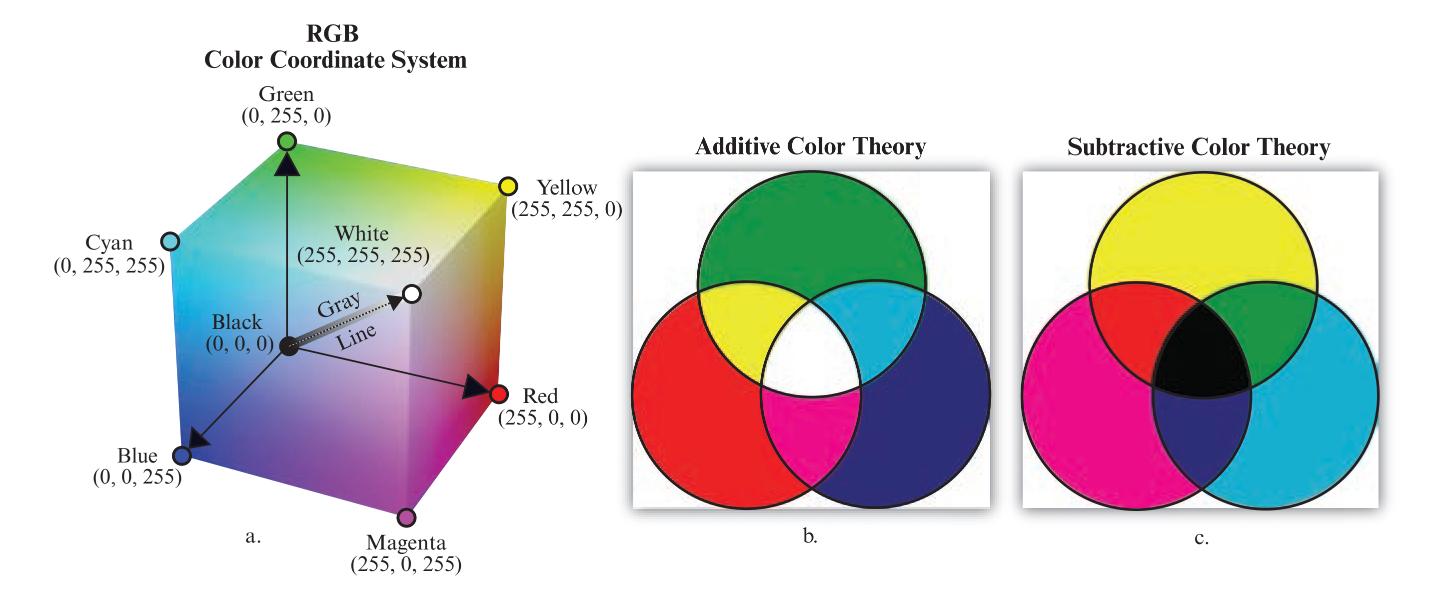


FIGURE 5–6 a) The RGB color coordinate system is based on additive color theory. If we are analyzing three 8-bit images (RGB) at one time, there is the possibility of 2^{24} = 16,777,216 unique values. Each of these unique values lies somewhere within the three-dimensional RGB color coordinate system. Black-and-white grayscale images are located along the diagonal from 0,0,0 to 255,255,255. b) ***Additive color theory*** — equal proportions of blue, green, and red light superimposed on top of one another creates white light, i.e., white light is composed of blue, green, and red light. The complementary colors yellow, magenta, and cyan are created by selectively adding together red and green, blue and red, and blue and green light, respectively. c) ***Subtractive color theory*** — equal proportions of blue, green, and red pigments yield a black surface. A yellow filter effectively absorbs all blue light, a magenta filter absorbs all green light, and a cyan filter absorbs all red light.

that whenever possible we should use at least 8-bit bit-mapped displays when analyzing 8-bit remotely sensed data.

The remainder of the discussion is based on the use of 8-bit imagery. However, numerous remote sensing systems now have 9-, 10-, 11-, and even 12-bit radiometric resolution per band. Therefore, it is important not to take this parameter for granted.

RGB Color Coordinate System

Digital remote sensor data are usually displayed using a **Red-Green-Blue (RGB) color coordinate system** (Figure 5-6a), which is based on additive color theory and the three primary colors of red, green, and blue (Figure 5-6b). **Additive color theory** is based on what happens when light is mixed, rather than when pigments are mixed using subtractive color theory. For example, in additive color theory a pixel having RGB values of 255, 255, 255 produces a bright white pixel (Figure 5-6b). Conversely, we would get a dark pigment if we mixed equally high proportions of blue, green, and red paint based on **subtractive color theory** as shown in Figure 5-6c. Using three 8-bit images and additive color theory, we can conceivably display 2^{24} = 16,777,216 color combinations (Lillesand et al., 2008). For example, RGB brightness values of 255, 255, 0 would yield a bright yellow pixel, and RGB brightness values of 255, 0, 0 would produce a bright red pixel. RGB values of 0, 0, 0 yield a black pixel. Grays are produced along the gray line in the RGB color coordinate system (Figure 5-6a) when equal proportions of blue, green, and red are encountered (e.g., an RGB of 127, 127, 127 produces a medium-gray pixel on the screen or hard-copy device).

Color Look-Up Tables: 8-Bit

How do we control the exact gray tone or color of the pixel on the computer screen after we have extracted a byte of remotely sensed data from the mass storage device? The gray tone or color of an individual pixel on a computer screen is controlled by the size and characteristics of a separate block of computer memory called a **color look-up table**, which contains the exact disposition of each combination of red, green, and blue values associated with each 8-bit pixel. Evaluating the nature of an 8-bit image processor and associated color look-up table (Figure 5-7) provides insight into the way the remote sensing brightness values and color look-up table interact. Two examples of color look-up tables are provided to demonstrate how black-and-white and color density-sliced brightness maps are produced.

In Example 1 (Figure 5-7), we see that the first 256 elements of the look-up table coincide with progressively greater values of the red, green, and blue components,

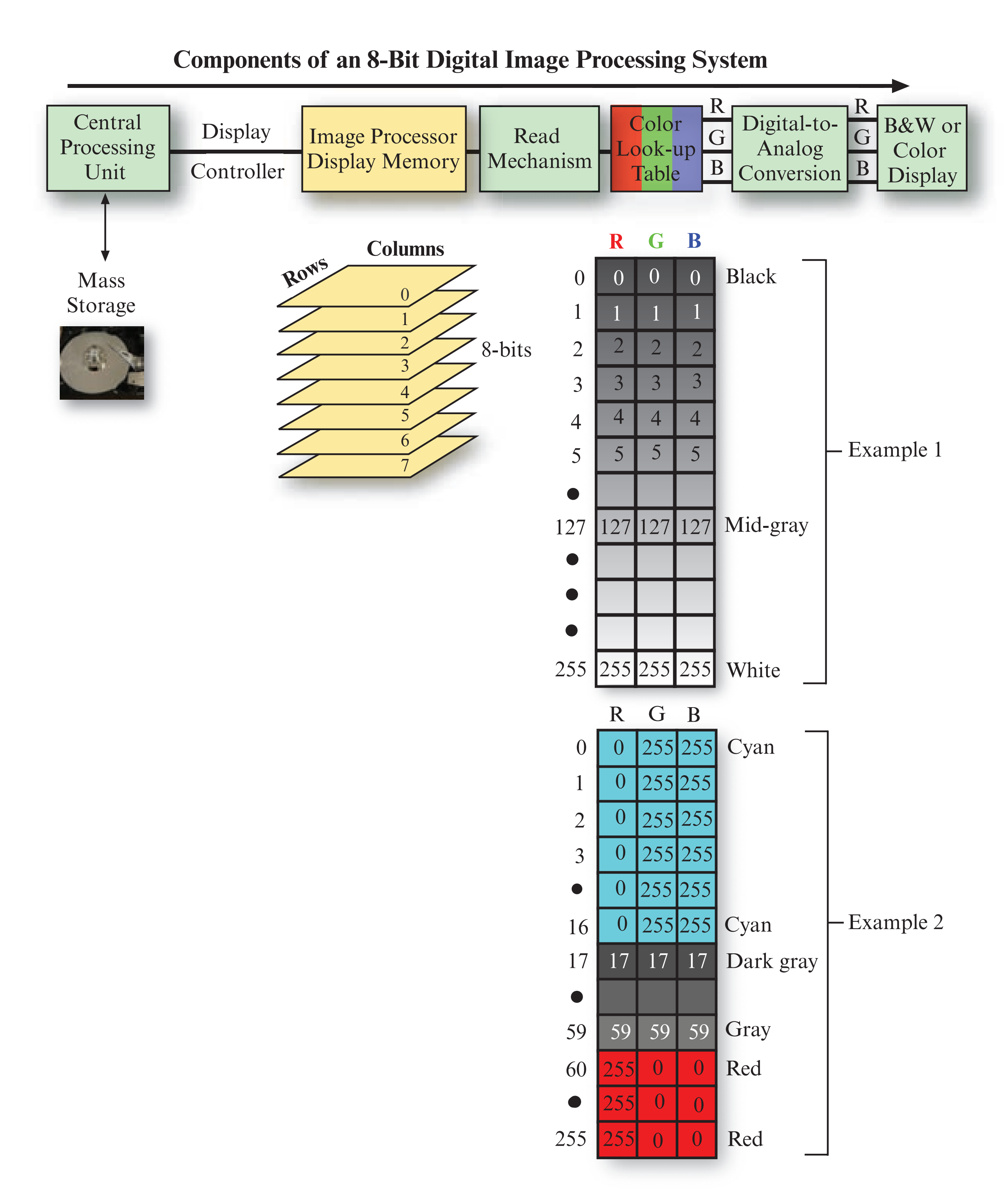
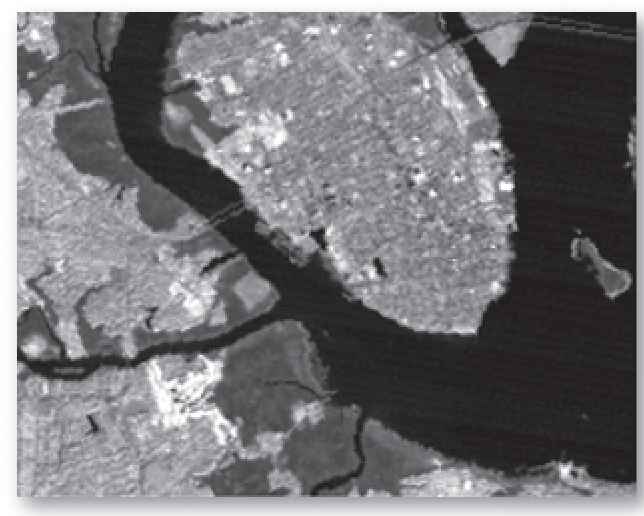


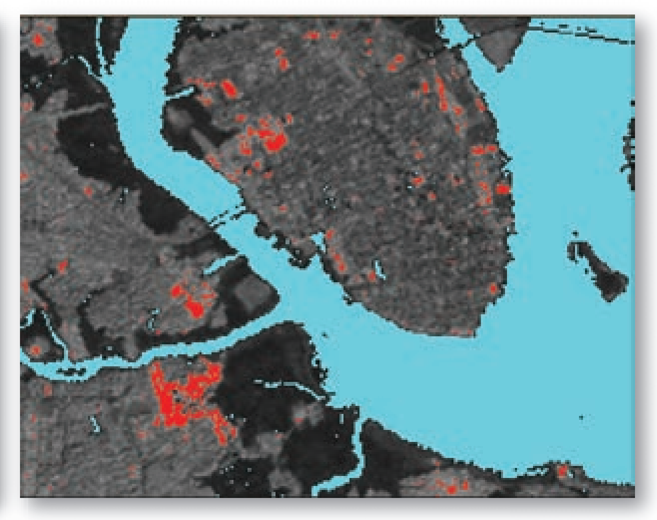
FIGURE 5–7 Components of an 8-bit digital image processing system. The image processor display memory is filled with the 8-bit remote sensing brightness values from a single band of imagery. These values are then manipulated for display by specifying the contents of the 256-element color look-up table. An 8-bit digital-to-analog converter (DAC) then converts the digital look-up table values into analog signals that are used to modulate the intensity of the red, green, and blue (RGB) guns that create the image on a color CRT screen. DACs are not required if a digital monitor is used. Also, most displays are now based on liquid crystal display (LCD) technology.

ranging from 0, 0, 0, which would be black on the screen, to 127, 127, 127, which would be mid-gray, and 255, 255, 255, which is bright white. Thus, if a pixel in a single band of remotely sensed data had a brightness value of 127, the RGB value located at table entry 127 (with RGB values of 127, 127, 127) would be passed to the 8-bit DAC converter and a mid-gray pixel would be displayed on the screen. This type of logic was applied

Density Slicing Using an 8-Bit Color Look-Up Table



a. Landsat Thematic Mapper band 4 image of Charleston, SC, obtained on November 9, 1982.

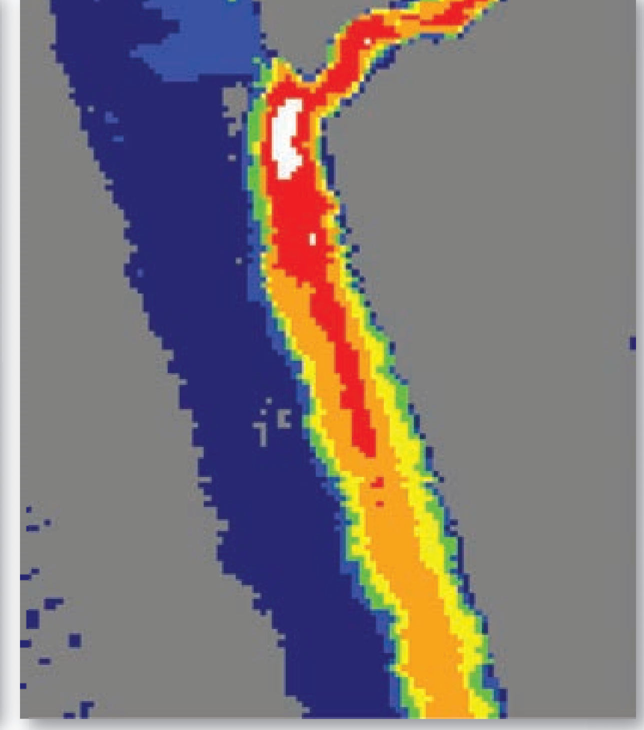


b. Density slice based on the logic in Table 5-2.





c. Predawn thermal infrared image of the Savannah River obtained on March 28, 1981.



d. Density slice based on the logic in Table 5-3.

FIGURE 5–8 a) Black-and-white display of Landsat Thematic Mapper band 4 (0.76 – 0.90 μm) 30 × 30 m data of Charleston, SC. b) Color density slice using the logic summarized in Table 5-2. c) Black-and-white display of predawn thermal infrared (8.5 – 13.5 μm) imagery of the Savannah River. Each pixel is approximately 2.8 × 2.8 m on the ground. d) Color density slice using the logic summarized in Table 5-3.

to create the black-and-white brightness map of Thematic Mapper band 4 data of Charleston, SC, shown in Figure 5-8a. This is often referred to as a true 8-bit black-and-white display because there is no generalization of the original remote sensor data. This means that there is a one-to-one relationship between the 8 bits of remote sensor input data and the 8-bit color look-up table (refer to Figure 5-2a). This is the ideal mechanism for displaying remote sensor data. The same logic was applied to produce the 8-bit display of

TABLE 5–2 Class intervals and color look-up table values for color density slicing the Charleston, SC, Thematic Mapper band 4 scene shown in Figure 5-8a and b.

Color Class Interval	Visual color	Color Look-up Table Value			Brightness Value	
		Red	Green	Blue	Low	High
1	Cyan	0	255	255	0	16
	Shade of gray	17	17	17	17	17
	Shade of gray	18	18	18	18	18
	Shade of gray	19	19	19	19	19
. . .	. . .	. . .	. . .	. . .	. . .	. . .
	Shade of gray	58	58	58	58	58
	Shade of gray	59	59	59	59	59
2	Red	255	0	0	60	255

TABLE 5–3 Class intervals and color look-up table values for color density slicing the predawn thermal infrared image of the thermal plume shown in Figure 5-8c and d.

Color Class Interval	Visual Color	Color Look-up Table Value			Apparent Temperature (°C)		Brightness Value	
		Red	Green	Blue	Low Value	Upper Value	Low	High
1. Land	Gray	127	127	127	–3.0	11.6	0	73
2. River ambient	Dark blue	0	0	120	11.8	12.2	74	76
3. +1°C	Blue	0	0	255	12.4	13.0	77	80
4. 1.2 – 2.8°C	Green	0	255	0	13.2	14.8	81	89
5. 3.0 – 5.0°C	Yellow	255	255	0	15.0	17.0	90	100
6. 5.2 – 10.0°C	Orange	255	50	0	17.2	22.0	101	125
7. 10.2 – 20.0°C	Red	255	0	0	22.2	32.0	126	176
8. >20°C	White	255	255	255	32.2	48.0	177	255

the predawn thermal infrared image of the Savannah River shown in Figure 5-8c.

Lay people and scientists who analyze remotely sensed data are rarely content with the display of the fundamental gray tone (individual band) or color (using multiple bands) information contained in a dataset. They usually want to highlight in color certain brightness values in the imagery associated with important phenomena (Richards, 2013; Konecny, 2014). The image analyst can accomplish this task by filling the color look-up table with very specific values (Tables 5-2 and 5-3).

To demonstrate this point, in Example 2 in Figure 5-7, the first entry in the table, 0, is given an RGB value of 0, 255, 255. Therefore, any pixel in a single band of imagery with a value of 0 would be displayed as cyan (a bright blue-green) on the computer screen. In this way it is possible to create a special-purpose look-up table with those colors in it that are of greatest value to the analyst. This is precisely the mechanism by which a single band of remote sensor data is color **density sliced**. For example, if we wanted to highlight just the water and the most healthy vegetation found within the Landsat TM band 4 image of Charleston, SC, we could density slice the image shown in Figure 5-8b based on the color look-up table values summarized in

Example 2 (Figure 5-7 and Table 5-2). The color look-up table has been modified so that all pixels between 0 and 16 have RGB values of 0, 255, 255 (cyan). Color look-up table values from 60 to 255 have RGB values of 255, 0, 0 (red). All values between 17 and 59 have the normal grayscale look-up table value; for example, a pixel with a BV of 17 has an RGB value 17, 17, 17, which will result in a dark gray pixel on the screen.

Entirely different brightness value class intervals were selected to density slice the predawn Savannah River thermal infrared image as shown in Figure 5-8d. The class intervals and associated color look-up table values for the Savannah River color density-sliced image are found in Table 5-3.

Color Look-Up Tables: 24-Bit

Much greater flexibility is provided when multiple 8-bit images can be stored and evaluated all at one time. For example, Figure 5-9 depicts the configuration of a 24-bit image processing system complete with three 8-bit banks of image processor memory and three 8-bit color look-up tables, one for each image memory bank. Thus, three separate 8-bit images could be stored at full resolution in the image processor memory banks (one in the red, one in the green, and one in the blue). Three separate 8-bit DACs continuously read the brightness value of a pixel in each of the red, green, and blue image planes and transform this digital value into an analog signal that can be used to modulate the intensity of the red, green, and blue (RGB) tricolor guns on the CRT screen. For example, if pixel (1, 1) in the red image plane has a brightness value of 255, and pixel (1, 1) in both the green and blue image planes have brightness values of 0, then a bright red pixel (255, 0, 0) would be displayed at location (1, 1) on the computer screen. More than 16.7 million RGB color combinations can be produced using this configuration. Obviously, this provides a much more ideal palette of colors to choose from, all of which can be displayed on the computer screen at one time. What we have just described is the fundamental basis behind the creation of additive color composites.

Color Composites

High spectral (color) resolution is important when producing color composites. For example, if a false-color reflective infrared TM image is to be displayed accurately, each 8-bit input image (TM band 4 = near infrared, TM band 3 = red, and TM band 2 = green) must be assigned 8 bits of red, green, and blue image processor memory, respectively. In this case, the 24-bit system with three 8-bit color look-up tables would provide a true-color rendition of each pixel on the screen, as previously discussed. This is true additive color combining with no generalization taking place.

Additive color composites produced from various Landsat TM band combinations of February 3, 1994, data are presented in Figure 5-10. The first is a natural color composite in which TM bands 3, 2, and 1 are placed in the red, green, and blue image processor display memory planes, respectively (Figure 5-10a). This is what the terrain would look like if the analyst were onboard the satellite platform looking down on South Carolina.

A color-infrared color composite of TM bands 4, 3, and 2 (RGB) is displayed in Figure 5-10b. Healthy vegetation shows up in shades of red because photosynthesizing vegetation absorbs most of the green and red incident energy but reflects approximately half of the incident near-infrared energy (discussed in Chapter 8). Dense urban areas reflect approximately equal proportions of near-infrared, red, and green energy; therefore they appear as steel gray. Moist wetland areas appear in shades of greenish brown. The third color composite was produced using bands 4 (near-infrared), 5 (middle-infrared), and 3 (red) (Figure 5-10c). The composite provides good definition of the land–water interface. Vegetation type and condition appear in shades of brown, green, and orange. The more moist the soil, the darker it appears. Figure 5-10d was created using TM bands 7, 4, and 2 (RGB). Many analysts like this combination because vegetation is presented in familiar green tones. Also, the mid-infrared TM band 7 helps discriminate moisture content in both vegetation and soils. Urban areas appear in varying shades of magenta. Dark green areas correspond to upland forest, while greenish brown areas are wetland. But what about all the other three-band color composites that can be produced from the same Landsat TM data?

Optimum Index Factor

Chavez et al. (1984) developed an **optimum index factor** (OIF) that ranks the 20 three-band combinations that can be made from six bands of TM data (not including the thermal-infrared band). The technique, however, is applicable to any multispectral remote sensing dataset. It is based on the amount of total variance and correlation within and between various band combinations. The algorithm used to compute the OIF for any subset of three bands is:

$$OIF = \frac{\sum_{k=1}^{3} s_k}{\sum_{j=1}^{3} \text{Abs}(r_j)}, \tag{5.1}$$

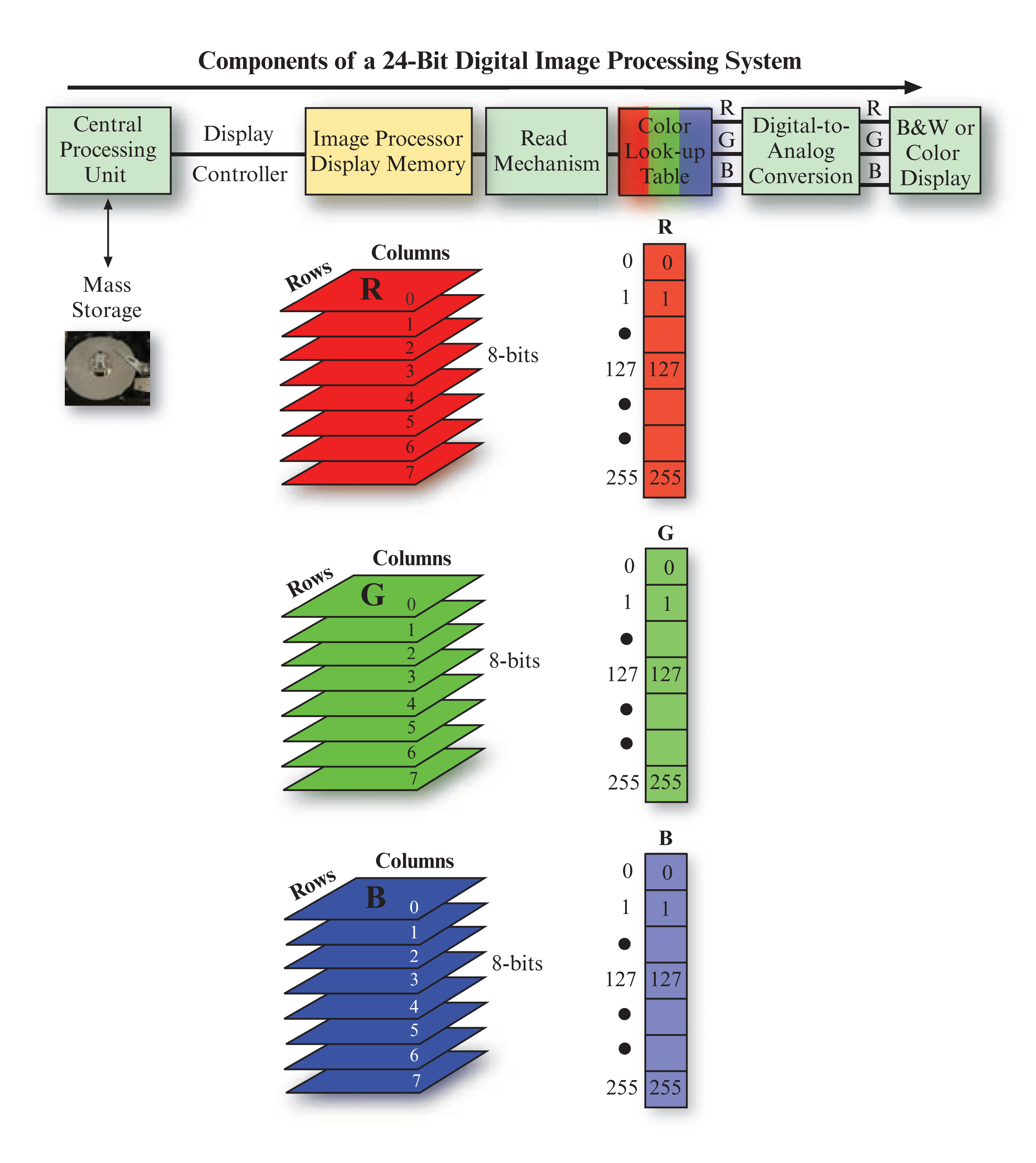


FIGURE 5–9 Components of a 24-bit digital image processing system. The image processor can store and continuously evaluate up to three 8-bit remotely sensed images. Three 8-bit digital-to-analog converters (DACs) scan the contents of the three 8-bit color look-up tables. The pixel color displayed on the screen will be just one possible combination (e.g., red = 255, 0, 0) out of a possible 16,777,216 colors that can be displayed at any time using the 24-bit image processing system. DACs are not required if a digital monitor is used.

where s_k is the standard deviation for band k, and r_j is the absolute value of the correlation coefficient between any two of the three bands being evaluated. The three-band combination with the largest OIF generally has the most information (measured by variance) with the least amount of duplication (measured by correlation). Combinations within two or three rankings of each other produce similar results.

Application of the OIF criteria to the 1982 Charleston, SC, Landsat TM dataset (excluding the thermal band) resulted in 20 combinations (Table 5-4). The standard

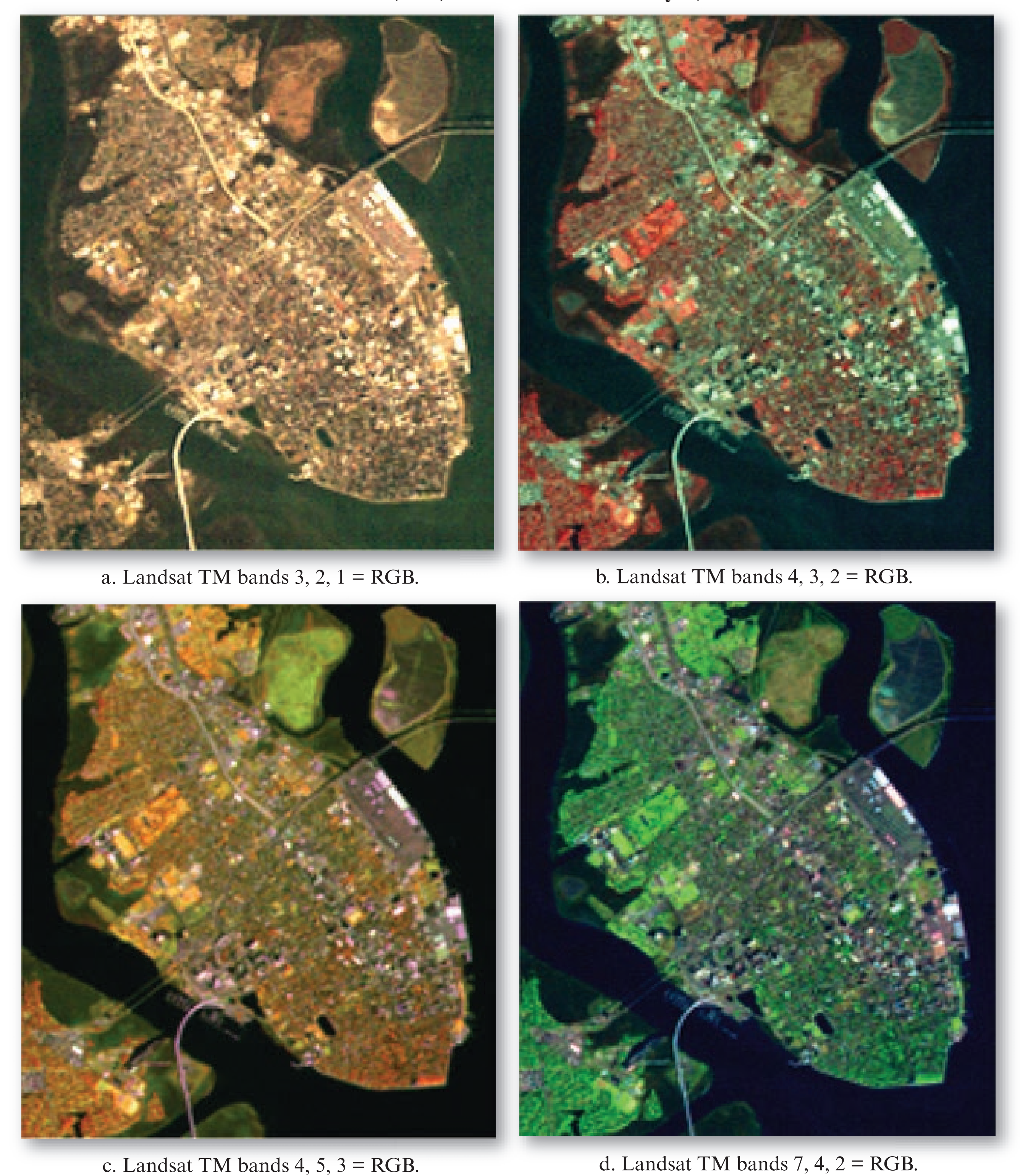


FIGURE 5–10 Color composites of Landsat Thematic Mapper data of Charleston, SC, collected on February 3, 1994. a) Composite of Landsat TM bands 3, 2, and 1 placed in the red, green, and blue (RGB) image processor memory planes, respectively. b) TM bands 4, 3, and 2 = RGB. c) TM bands 4, 5, 3 = RGB. d) TM bands 7, 4, 2 = RGB. Original imagery courtesy of NASA.

TABLE 5–4 Optimum Index Factors for six of the Charleston, SC, Landsat Thematic Mapper bands.

Rank	Combination	OIF
1	1, 4, 5	27.137
2	2, 4, 5	23.549
3	3, 4, 5	22.599
4	1, 5, 7	20.785
5	1, 4, 7	20.460
6	4, 5, 7	20.100
7	1, 3, 5	19.009
8	1, 2, 5	18.657
9	1, 3, 4	18.241
10	2, 5, 7	18.164
11	2, 3, 5	17.920
12	3, 5, 7	17.840
13	1, 2, 4	17.682
14	2, 4, 7	17.372
15	3, 4, 7	17.141
16	2, 3, 4	15.492
17	1, 3, 7	12.271
18	1, 2, 7	11.615
19	2, 3, 7	10.387
20	1, 2, 3	8.428

Six bands combined three at a time allows 20 combinations. The thermal infrared band 6 (10.4 to 12.5 μm) was not used.

For example, the OIF for Landsat TM band combination 1, 4, and 5 from Table 4-7 is:

$$\frac{10.5 + 15.76 + 23.85}{0.39 + 0.56 + 0.88} = 27.137 .$$

deviations and between-band correlation coefficients were obtained from Table 4-7. A three-band combination using bands 1, 4, and 5 should provide the optimum color composite, with bands 2, 4, and 5 and 3, 4, and 5 just about as good. Generally, the best three-band combinations include one of the visible bands (TM 1, 2, or 3), and one of the longer-wavelength infrared bands (TM 5 or 7), along with TM band 4. TM band 4 was present in five of the first six rankings. Such information can be used to select the most useful bands for three-band color composites. The analyst must then decide what color to assign each band (red, green, or blue) in the color composite.

Sheffield Index

Sheffield (1985) developed a statistical band selection index based on the size of the hyperspace spanned by the three bands under investigation. Sheffield suggested that the bands with the largest hypervolumes be selected. The index is based on computing the determinant of each p by p sub-matrix generated from the original 6 × 6 covariance matrix (if six bands are under investigation). The *Sheffield Index* (*SI*) is:

$$SI = \left|Cov_{p \times p}\right| , \tag{5.2}$$

where $\left|Cov_{p \times p}\right|$ is the determinant of the covariance matrix of subset size p. In this case, $p = 3$ because we are trying to discover the optimum three-band combination for image display purposes. In effect, the SI is first computed from a 3 × 3 covariance matrix derived from just band 1, 2, and 3 data. It is then computed from a covariance matrix derived from just band 1, 2, and 4 data, etc. This process continues for all 20 possible band combinations if six bands are under investigation, as in the previous example. The band combination that results in the largest determinant is selected for image display. All of the information necessary to compute the SI is actually present in the original 6 × 6 covariance matrix. The Sheffield Index can be extended to datasets containing n bands. Beauchemin and Fung (2001) suggested using a normalized version of the Sheffield Index.

Independent Component Analysis-Based (ICA-Based) Fusion for Color Display of Hyperspectral Images

Hyperspectral imagery often consists of many hundreds of bands. It is difficult to know which three of the bands to display on a computer screen. Zhu et al. (2011) developed an ICA-based dimensionality reduction method to fuse the n hyperspectral images to three independent component images for color display. They suggest it performs better than standard principal components analysis (PCA) dimensionality reduction for hyperspectral data display.



Merging (Fusing) Remotely Sensed Data

Image analysts often merge (fuse) different types of remote sensor data such as:

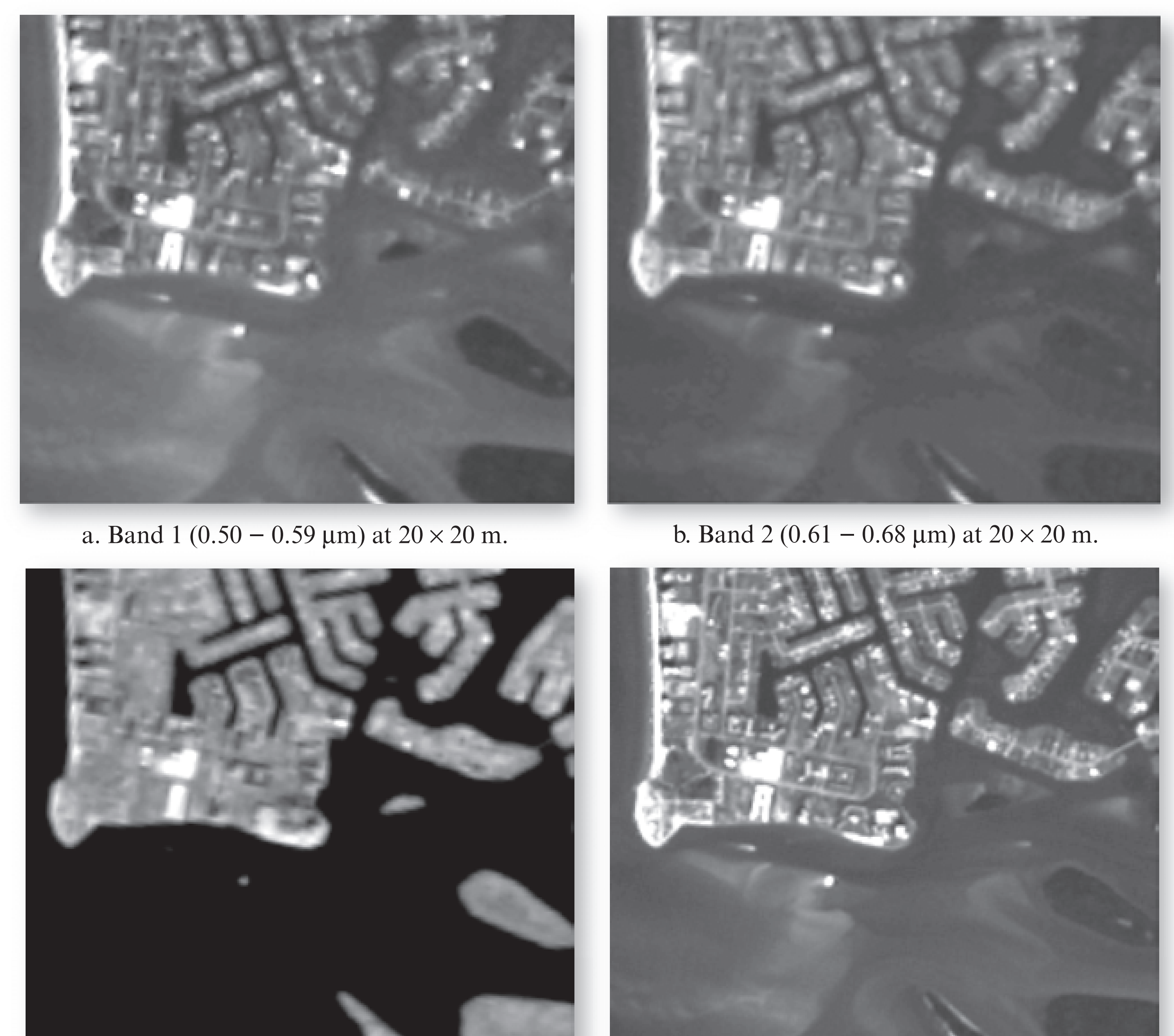


FIGURE 5–11 Individual bands of SPOT 20 × 20 multispectral and 10 × 10 m panchromatic imagery used to create the normal color-infrared color composites in Figure 5-12a and the merged (fused) color-infrared color composite in Figure 5-12b [images courtesy of CNES (www.cnes.fr). Protected information. All rights reserved © CNES (2014)].

- SPOT 10 × 10 m PAN data with Landsat Thematic Mapper 30 × 30 m data (e.g., Chavez and Bowell, 1988),
- Multispectral data (e.g., SPOT XS, Landsat TM, IKONOS) with active microwave (radar) and/or other data (e.g., Chen et al., 2003; Klonus and Ehlers, 2009),
- IKONOS 1 × 1 m data with Landsat Enhanced Thematic Mapper (ETM^+) 30 × 30 m data (Chen et al., 2011),
- Landsat ETM^+ multispectral 30 × 30 m data with ETM^+ 15 × 15 m panchromatic data (Park and Kang, 2004),
- QuickBird 0.7 × 0.7 m panchromatic with 2.8 × 2.8 m multispectral data (Aiazzi et al., 2007),
- Hyperspectral data into three independent components for visual display (Zhu et al., 2011), and
- CARTOSAT-1 panchromatic 2.5 × 2.5 m data and IRS-P6 LISS-IV 5.8 × 5.8 m multispectral data (Jalan and Sokhi, 2012).

When lower spatial resolution imagery such as 30 × 30 m Landsat Thematic Mapper data are fused (merged) with higher spatial resolution panchromatic imagery such as 1 × 1 m IKONOS panchromatic data, the process is commonly referred to as **pan-sharpening**. The goal of pan-sharpening is to render a sharpened image incorporating the full spatial information content of the panchromatic image without introducing spectral distortions to the multispectral input data (Jalan and Sokhi, 2012).

Merging (fusing) remotely sensed data obtained using different remote sensors must be performed carefully. All datasets to be merged must be accurately registered to one another and resampled (Chapter 6) to the same pixel size. Several component-substitution (CS) methods exist for merging (fusing) the data sets, including (Park and Kang, 2004; Schowengerdt, 2006; Aiazzi et al., 2007; Alparone et al., 2007; Klonus and Ehlers, 2009; Prost, 2013):

- simple band-substitution methods;
- color space transformation and component substitution methods using RGB, intensity–hue–saturation, and chromaticity color coordinate systems;
- Principal Component Analysis (PCA), Independent Component Analysis (ICA) or Gram-Schmidt substitution;
- pixel-by-pixel addition of high-frequency information to a lower spatial resolution dataset;
- image fusion based on regression kriging; and
- smoothing filter-based intensity modulation image fusion.

There are other pan-sharpening algorithms available such as multi-resolution analysis (MRA) based on the use of wavelets or Laplacian pyramids and Ehler's fusion method based on an IHS transform and Fourier filtering that are not discussed in this introductory text (e.g., Alparone et al., 2007; Klonus and Ehlers, 2007).

Simple Band Substitution

Simple band substitution is one of the most heavily used image fusion methods. For example, individual SPOT bands of Marco Island, FL, are displayed in Figure 5-11. The data were geometrically rectified to a Universal Transverse Mercator (UTM) projection and then resampled to 10 × 10 m pixels using bilinear interpolation. Note how the 20 × 20 m multispectral bands (green, red, and near-infrared) appear out of focus. This is typical of 20 × 20 m data of urban areas. Conversely, the 10 × 10 m panchromatic data exhibits more detailed road network information and perimeter detail for some of the larger buildings. Merging (fusion) of the 20 × 20 m multispectral data with the 10 × 10 m panchromatic data could be of value.

Figure 5-12a is a color-infrared color composite with the band 3 (near-infrared), band 2 (red), and band 1 (green) imagery placed in the red, green, and blue image processor display memory banks, respectively (i.e., bands 3, 2, 1 = RGB). The SPOT panchromatic data span the spectral region from 0.51 to 0.73 μm. Therefore, it is a record of both green and red energy. It can be substituted directly for either the green (SPOT 1) or red (SPOT 2) bands. Figure 5-12b is a display of the merged (fused) dataset with SPOT 3 (near-infrared), SPOT panchromatic, and SPOT 1 (green) in the RGB image processor memory planes, respectively. The result is a display that contains the spatial detail of the SPOT panchromatic data (10 × 10 m) and spectral detail of the 20 × 20 m SPOT multispectral data. This method has the advantage of not changing the radiometric qualities of any of the SPOT data.

Color Space Transformation and Component Substitution

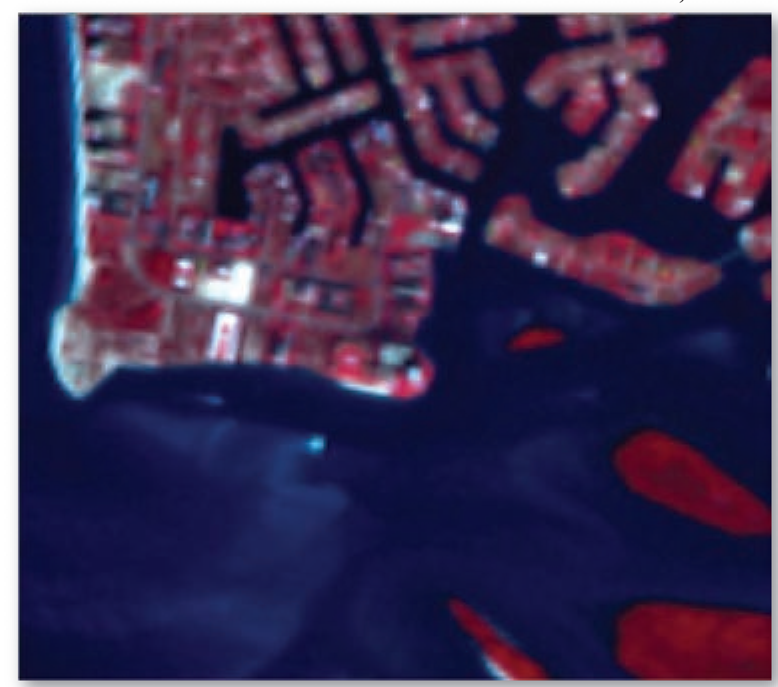
All remotely sensed data presented thus far have been in the RGB color coordinate system. Other color coordinate systems may be of value when presenting remotely sensed data for visual analysis, and some of these may be used when different types of remotely sensed data are merged. Two frequently used methods are the RGB to intensity–hue–saturation (IHS) transformation and the use of chromaticity coordinates.

RGB to IHS Transformation and Back Again

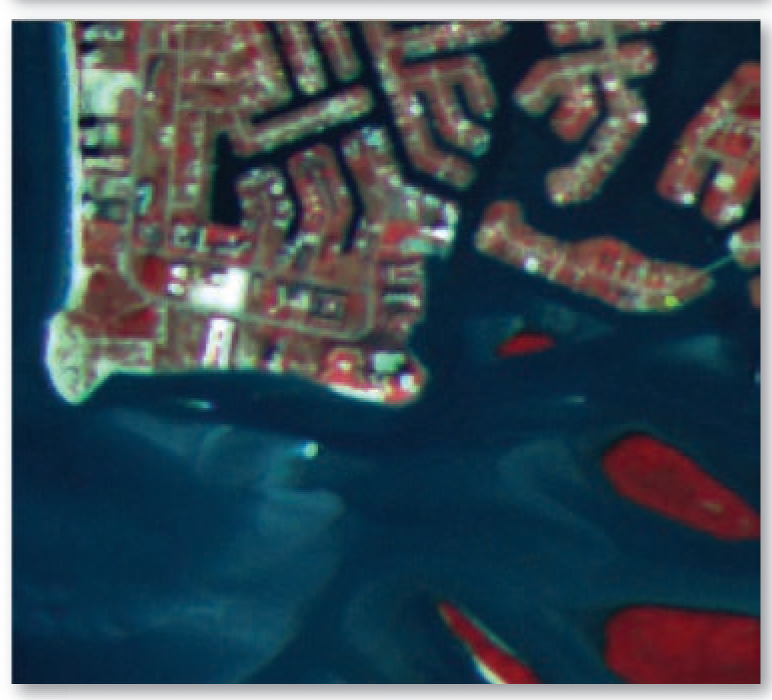
The **intensity–hue–saturation (IHS) color coordinate system** is based on a hypothetical color sphere (Figure 5-13a). The vertical axis represents *intensity* (I) which varies from black (0) to white (255) and is not associated with any color. The circumference of the sphere represents *hue* (H), which is the dominant wavelength of color. Hue values begin with 0 at the midpoint of red tones and increase counterclockwise around the circumference of the sphere to conclude with 255 adjacent to 0. *Saturation* (S) represents the purity of the color and ranges from 0 at the center of the color sphere to 255 at the circumference. A saturation of 0 represents a completely impure color in which all wavelengths are equally represented and which the eye will perceive as a shade of gray that ranges from white to black depending on intensity. Intermediate values of saturation represent pastel shades, whereas high values represent more pure, intense colors. All values used in this example are scaled to 8 bits, corresponding to most digital remote sensor data. Figure 5-13a highlights the location of a single pixel with IHS coordinates of 190, 0, 220 (i.e., a relatively intense [190], highly saturated [220], red [0] pixel).

Any RGB multispectral dataset consisting of three bands may be transformed into IHS color coordinate

Merging (Fusion) of SPOT 20 × 20 m Multispectral and 10 × 10 m Panchromatic Data of Marco Island, FL



a. Color-infrared color composite of SPOT band 3 (near-infrared), 2 (red), and 1 (green) = RGB. Each band is 20 × 20 m.



b. Fused color-infrared color composite of SPOT band 3 (near-infrared), 4 (panchromatic), and 1 (green) = RGB. The panchromatic band is 10 × 10 m. The composite was created by substituting the panchromatic band for band 2 (red).

FIGURE 5–12 a) Merging (fusion) of SPOT multispectral data (20 × 20 m) with SPOT panchromatic data (10 × 10 m) using the band substitution method. The 20 × 20 m multispectral data were resampled to 10 × 10 m [images courtesy of CNES (www.cnes.fr). Protected information. All rights reserved © CNES (2014)].

space using an IHS transformation. This transformation is actually a limitation because many remote sensing datasets contain more than three bands. The relationship between the RGB and IHS systems is shown diagrammatically in Figure 5-13b. Numerical values may be extracted from this diagram for expressing either system in terms of the other. The circle represents a horizontal section through the equatorial plane of the IHS sphere (Figure 5-13a) with the intensity axis passing vertically through the plane of the diagram. The corners of the equilateral triangle are located at the position of the red, green, and blue hues. Hue

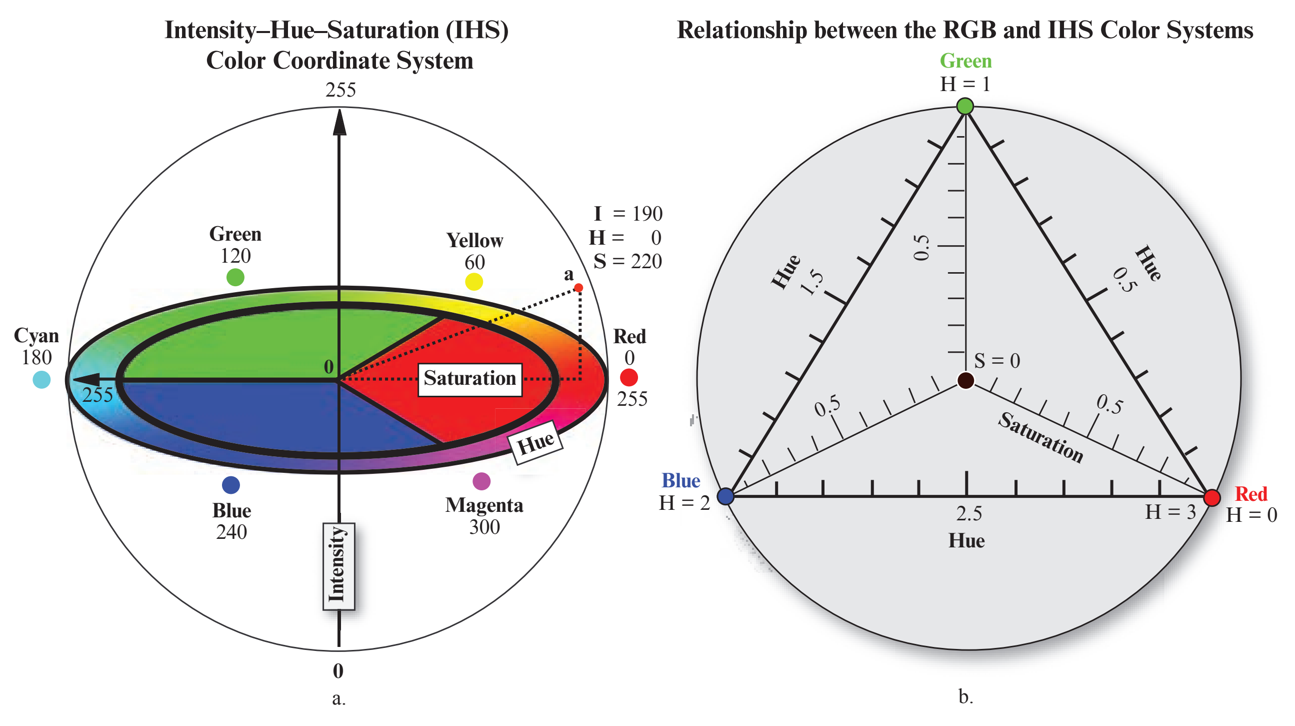


FIGURE 5–13 a) Intensity–hue–saturation (IHS) color coordinate system. The primary colors are on the interior of the circle. The secondary colors are on the periphery of the circle. b) Relationship between the intensity–hue–saturation (IHS) color coordinate system and the RGB color coordinate system.

changes in a counterclockwise direction around the triangle, from red (H = 0), to green (H = 1), to blue (H = 2), and again to red (H = 3). Values of saturation are 0 at the center of the triangle and increase to a maximum of 1 at the corners. Any perceived color is described by a unique set of IHS values. The IHS values can be derived from the RGB values through transformation equations (Sabins, 2007):

$$I = R + G + B, \quad (5.3)$$

$$H = \frac{G - B}{I - 3B}, \quad (5.4)$$

and

$$S = \frac{I - 3B}{I} \quad (5.5)$$

for the interval 0 < H < 1, extended to 1 < H < 3. Pellemans et al. (1993) used different equations to compute intensity, hue, and saturation for a SPOT dataset consisting of three bands of remotely sensed data (BV_1, BV_2, and BV_3):

$$Intensity = \frac{BV_1 + BV_2 + BV_3}{3}, \quad (5.6)$$

$$Hue = \arctan\frac{2BV_1 - BV_2 - BV_3}{\sqrt{3}(BV_2 - BV_3)} + C \quad (5.7)$$

$$\text{where} \begin{cases} C = 0, \text{ if } BV_2 \geq BV_3 \\ C = \pi, \text{ if } BV_2 < BV_3 \end{cases},$$

and

$$Saturation =$$

$$\frac{\sqrt{6(BV_1^2 + BV_2^2 + BV_3^2 - BV_1BV_2 - BV_1BV_3 - BV_2BV_3)}^{-0.5}}{3}. \quad (5.8)$$

So what is the benefit of performing an IHS transformation? First, it may be used to improve the interpretability of multispectral color composites. When any three spectral bands of multispectral data are combined in the RGB system, the color composite image often lacks saturation, even when the bands have been contrast stretched. Therefore, some analysts perform an RGB-to-IHS transformation, contrast stretch the resultant saturation image, and then convert the IHS images back into RGB images using the inverse of the equations just presented. The result is usually an improved color composite.

The IHS transformation is also often used to merge (fuse) lower spatial resolution imagery such as 30 × 30 m Landsat TM data with higher spatial resolution data such as IKONOS panchromatic 1 × 1 m imagery (e.g., Gonzalez-Audicana, 2004; Park and Kang, 2004;

Schowengerdt, 2006; Aiazzi et al., 2007; Yao and Han, 2010). The method generally involves four steps:

1. *RGB to IHS:* Three bands of lower-spatial-resolution remote sensor data in RGB color space are transformed into three bands in IHS color space.
2. *Contrast manipulation:* The high-spatial-resolution image (e.g., SPOT PAN data or digitized aerial photography) is contrast stretched so that it has approximately the same variance and mean as the intensity (I) image.
3. *Substitution:* The stretched, high-spatial-resolution image is substituted for the intensity (I) image.
4. *IHS to RGB:* The modified IHS dataset is transformed back into RGB color space using an inverse IHS transformation. The justification for replacing the intensity (I) component with the stretched higher-spatial-resolution image is that the two images have approximately the same *spectral* characteristics

Ehlers et al. (1990) used this methodology to merge SPOT 20 × 20 m multispectral and SPOT panchromatic 10 × 10 m data. The resulting multiresolution image retained the spatial resolution of the 10 × 10 m SPOT panchromatic data, yet provided the spectral characteristics (hue and saturation values) of the SPOT multispectral data. The enhanced detail available from merged images was found to be important for visual land-use interpretation and urban growth delineation (Ehlers et al., 1990). In a similar study, Carper et al. (1990) found that direct substitution of the panchromatic data for intensity (I) derived from the multispectral data was not ideal for visual interpretation of agricultural, forested, or heavily vegetated areas. They suggested that the original intensity value obtained in step 1 be computed using a weighted average (WA) of the SPOT panchromatic and SPOT multispectral data; that is, WA = {[(2 × SPOT Pan) + SPOT XS3]/3}.

Chavez et al. (1991) cautioned that of all the methods used to merge multi-resolution data, the IHS method distorts the spectral characteristics the most and should be used with caution if detailed radiometric analysis of the data is to be performed. Fortunately, numerous scientists have improved upon the IHS substitution method. For example, Park and Kang (2004) developed a spatially adaptive image fusion algorithm that controls the amount of high frequency information extracted from the high spatial resolution image that is inserted into the resampled multispectral data. Choi (2006) introduced an IHS fusion pan-sharpening method. Aiazzi et al. (2007) used multivariate regression to retain the quality of the spectral information without sacrificing spatial detail.

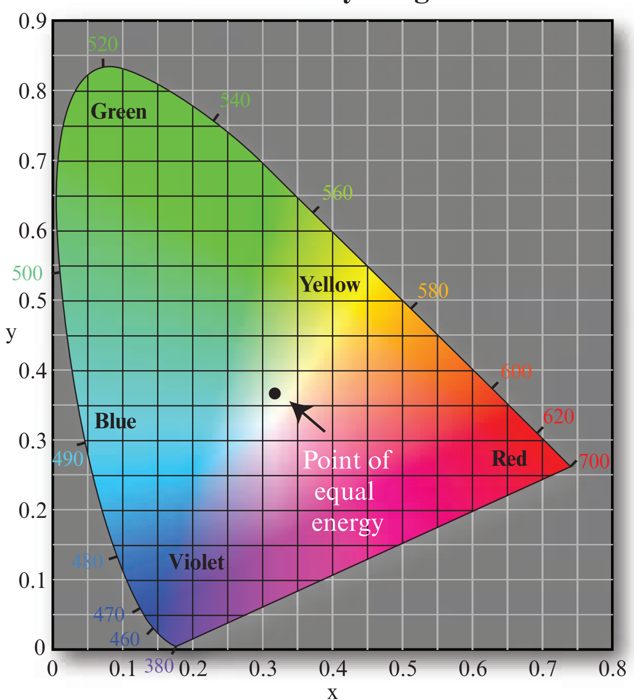


FIGURE 5–14 Chromaticity color coordinate system.

Chromaticity Color Coordinate System and the Brovey Transformation

A **chromaticity color coordinate system** can be used to specify color (Ready, 2013). A chromaticity diagram is shown in Figure 5-14. The coordinates in the chromaticity diagram represent the relative fractions of each of the primary colors (red, green, and blue) present in a given color. Since the sum of all three primaries must add to 1, we have the relationship:

$$R + G + B = 1 \tag{5.9}$$

or

$$B = 1 - (R + G). \tag{5.10}$$

Entry into the chromaticity diagram is made using the following relationships:

$$x = \frac{R}{R+G+B}, \tag{5.11}$$

$$y = \frac{G}{R+G+B}, \tag{5.12}$$

and

$$z = \frac{B}{R+G+B}, \tag{5.13}$$

where R, G, and B represent the amounts of red, green, and blue needed to form any particular color, and x, y, and z represent the corresponding *normalized* color

components, also known as *trichromatic coefficients* (Kulkarni, 2001). Only x and y are required to specify the chromaticity coordinates of a color in the diagram since $x + y + z = 1$. For example, the point marked "green" in Figure 5-14 has 70% green and 10% red content. It follows from Equation 5.10 that the composition of blue is 20%.

The human eye responds to light only in the range 380 to 800 nm (Myler, 2013). Therefore, the positions of the various spectrum colors from 380 to 800 nm are indicated around the boundary of the chromaticity diagram. These are the pure colors. Any point not actually on the boundary but within the diagram represents some mixture of spectrum colors. The point of equal energy shown in Figure 5-14 corresponds to equal fractions of the three primary colors and represents the CIE standard for white light. Any point located on the boundary of the chromaticity diagram is said to be completely saturated. As a point leaves the boundary and approaches the point of equal energy, more white light is added to the color and it becomes less saturated. The chromaticity diagram is useful for color mixing, because a straight-line segment joining any two points in the diagram defines all of the color variations that can be formed by combining the *two* colors additively.

The **Brovey transform** may be used to merge (fuse) images with different spatial and spectral characteristics. It is based on the chromaticity transform and is a much simpler technique than the RGB-to-IHS transformation. The Brovey transform also can be applied to individual bands if desired. It is based on the following intensity modulation (Liu, 2000a):

$$Red_{Brovey} = \frac{R \times P}{I}, \tag{5.14}$$

$$Green_{Brovey} = \frac{G \times P}{I}, \tag{5.15}$$

$$Blue_{Brovey} = \frac{B \times P}{I}, \tag{5.16}$$

and

$$I = \frac{R + G + B}{3}, \tag{5.17}$$

where R, G, and B are the spectral band images of interest (e.g., 30 × 30 m Landsat ETM$^+$ bands 4, 3, and 2) to be placed in the red, green, and blue image processor memory planes, respectively, P is a coregistered band of higher spatial resolution data (e.g., 1 × 1 m IKONOS panchromatic data), and I = intensity. An example of the Brovey transform used to merge IKONOS 4 × 4 m multispectral data (bands 4, 3, and 2) with IKONOS 1 × 1 m panchromatic data is shown in Figure 5-15. The merged (fused) dataset now has the spectral characteristics of the multispectral data and the spatial characteristics of the high-resolution panchromatic data.

Both the RGB-to-IHS transformation and the Brovey transform can cause color distortion if the spectral range of the intensity replacement (or modulation) image (i.e., the panchromatic band) is different from the spectral range of the three lower-resolution bands. The Brovey transform was developed to visually increase contrast in the low and high ends of an image's histogram (i.e., to provide contrast in shadows, water, and high-reflectance areas such as urban features). Consequently, the Brovey transform should not be used if preserving the original scene radiometry is important. However, it is good for producing RGB images with a higher degree of contrast in the low and high ends of the image histogram and for producing visually appealing images.

Principal Component Analysis (PCA), Independent Component Analysis (ICA), or Gram-Schmidt Substitution

PCA is a commonly used method for spectral transformation of the original remote sensor data to produce uncorrelated principal component images (Pratt, 2013) (PCA is discussed in detail in Chapter 8). PCA-based pan-sharpening methods make the assumption that the first principal component image (e.g., PC1) explains the largest amount of variance in the entire original remote sensing dataset (Shah et al., 2008). Therefore, PC1 is an excellent choice to be replaced with the high spatial resolution image, usually a panchromatic image. An inverse principal component analysis is then applied to the substituted PC1 image along with the other principal component images to obtain a high-resolution pan-sharpened image. The panchromatic image is usually histogram-matched with the first PC1 image before the substitution takes place (Campbell, 2007). For example, Chavez et al. (1991) used PCA applied to six 30 × 30 m Landsat TM bands. The SPOT panchromatic 10 × 10 m data were contrast stretched to have approximately the same variance and average as the first principal component image (PC1). The stretched panchromatic data were substituted for the first principal component image and the data were transformed back into RGB space. The advantage of PCA-based pan sharpening is that the number of bands is not restricted (such as for the original IHS or Brovey fusion methods) (Klonus and Ehlers, 2009). Shah et al. (2008) and Yang and Gong (2012) developed improved PCA substitution pan-sharpening methods.

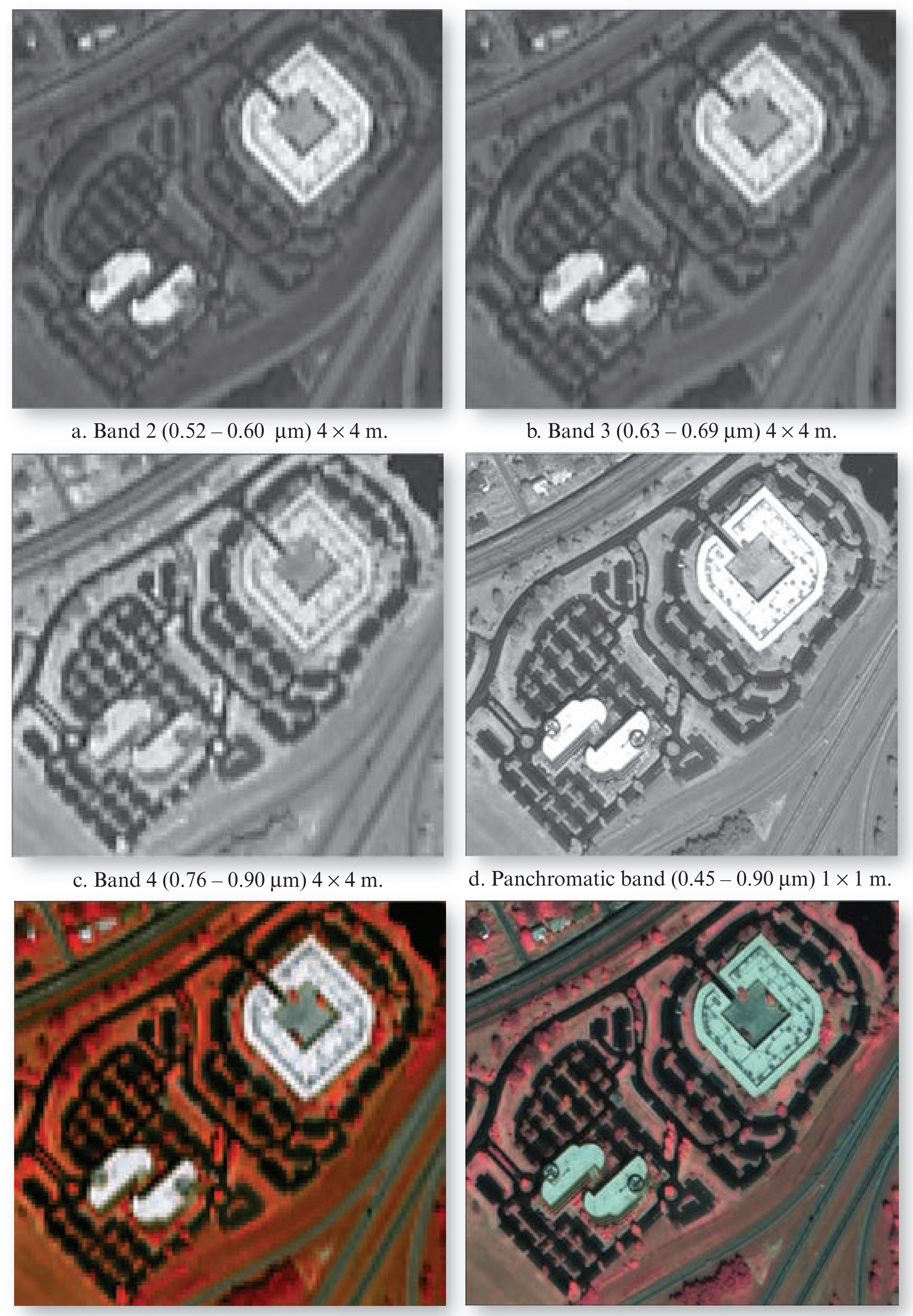


FIGURE 5–15 IKONOS imagery of a business park in Columbia, SC. a-d) Individual 4 × 4 m multispectral bands and the 1 × 1 m panchromatic band are displayed. e) Standard color composite of IKONOS bands 4, 3, and 2. f) Color composite of a merged dataset created using a Brovey transform (images courtesy of DigitalGlobe, Inc.).

Chen et al. (2011) used independent component analysis (ICA) to separate the main body, spectral detail, and spatial detail in a Landsat ETM$^+$ 30 × 30 m multispectral dataset. They then substituted the information from an IKONOS high spatial resolution 1 × 1 m panchromatic image for the main body independent component data. After inversion back to RGB space, the new data (image) contained the spatial information of the high-resolution image and the spectral characteristics of the original multispectral image.

The **Gram-Schmidt image fusion** method first synthesizes a panchromatic band from the lower spatial resolution spectral bands (Klonus and Ehlers, 2009). This is achieved by averaging the multispectral bands. A Gram-Schmidt transformation is then performed for the simulated panchromatic band and the multispectral bands with the simulated panchromatic band employed as the first band in the dataset. Then the high spatial resolution panchromatic band replaces the first Gram-Schmidt band. Finally, an inverse Gram-Schmidt transformation is performed to create the pan-sharpened multispectral bands in RGB color space (Laban et al., 2000). Klonus and Ehler (2009) suggest that the method produces good results for fusing images from one sensor, but because it is a statistical procedure like principal components analysis (PCA), the fusion results may vary depending on the selected image subsets. Gram-Schmidt image fusion is available in ENVI software.

Pixel-by-Pixel Addition of High-Frequency Information

Schowengerdt (1980) introduced the concept of utilizing image-derived high-frequency information in the fusion process. Chavez (1986), Chavez and Bowell (1988), and Chavez et al. (1991) merged both digitized National High Altitude Program photography and SPOT panchromatic data with Landsat Thematic Mapper data using a high-pass spatial filter applied to the high-spatial-resolution imagery. The resultant high-pass image contains high-frequency information that is related mostly to spatial characteristics of the scene. The spatial filter removes most of the spectral information. The high-pass filter results were added, pixel by pixel, to the lower-spatial-resolution TM data. This process merged the spatial information of the higher-spatial-resolution dataset with the higher spectral resolution inherent in the TM dataset. Chavez et al. (1991) found that this multisensor fusion technique distorted the spectral characteristics the least.

Gonzalez-Audicana et al. (2004) improved upon the use of specially-prepared high-pass filtered data by using wavelet decomposition. Jalan and Sokhi (2012) compared the high-pass filtering (HPF) pan-sharpening fusion method with four other image fusion methods. The HPF method was capable of providing fusion results with high spectral integrity as well as good spatial enhancement when applied to CARTOSAT-1 panchromatic data (2.5 × 2.5 m) and IRS-P6 LISS-IV multispectral imagery (5.8 × 5.8 m).

Fusion based on Regression Kriging

Geostatistics and kriging were introduced in Chapter 4. Meng et al. (2010) developed an image fusion method based on regression kriging which takes into consideration the correlation between the image to be fused and the image with finer spatial resolution, the spatial autocorrelation among pixels in the high spatial resolution image, and the unbiased estimation with minimized variance. Qualitative assessments suggest that there is no apparent color distortion in the fused images. Van der Meer (2012) reviews other geostatistical image fusion research.

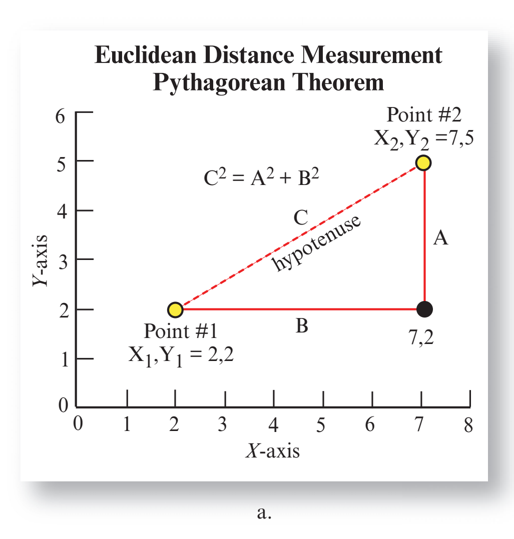
Smoothing Filter-Based Intensity Modulation Image Fusion

Liu (2000ab) developed a *Smoothing Filter-based Intensity Modulation* (SFIM) image fusion technique based on the algorithm:

$$BV_{SFIM} = \frac{BV_{low} \times BV_{high}}{BV_{mean}} \tag{5.18}$$

where BV_{low} is a pixel from the low spatial resolution image, BV_{high} is a pixel from the high spatial resolution image, and BV_{mean} is a simulated low resolution pixel derived from the high-resolution image using an averaging filter for a neighborhood equivalent in size to the spatial resolution of the low-resolution data. For example, suppose the high-resolution image consisted of SPOT 10 × 10 m panchromatic data and the low-resolution image consisted of Landsat ETM$^+$ 30 × 30 m data. In this case, the BV_{mean} value would be the average of the nine 10 × 10 m pixels centered on the pixel under investigation in the high spatial resolution dataset. Liu (2000a) suggests that the SFIM can produce optimally fused data without altering the spectral properties of the original image if the coregistration error is minimal.

Liu et al. (2012) undertook an objective assessment of various multiresolution image fusion algorithms. Khaleghi et al. (2013) provided a taxonomy of data fusion methodologies.



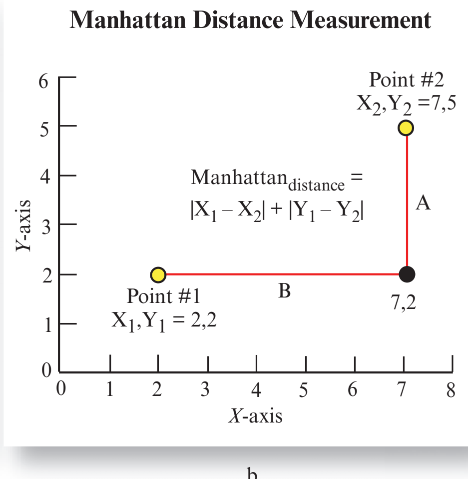


FIGURE 5–16 a) The calculation of the Euclidean distance between two points in an *X,Y* Cartesian coordinate system using the Pythagorean theorem. b) The calculation of the Manhattan distance between Point #1 and Point #2 in an *X, Y* Cartesian coordinate system.



Length (Distance) Measurement

Sometimes while viewing an image on the computer screen, an analyst desires to measure the length of an object, its area, or its shape. Most digital image processing programs have graphical user interfaces that allow the analyst to first identify the point, line, or area feature in the imagery and then extract the type of measurement desired.

Linear Distance Measurement Based on the Pythagorean Theorem

One of the most common measurements you will need to calculate is the **Euclidean distance** between two points, which is easily computed using the Pythagorean theorem. The Pythagorean theorem is based on the relationship between the three sides of a right triangle and is typically stated as follows: In any right triangle (a triangle where one inside angle = 90°), the length of the line segment opposite the right angle (the hypotenuse) is equal to the square root of the squared sums of the lengths of the other two legs (i.e., the two sides other than the hypotenuse). These relationships are shown in Figure 5-16a.

In mathematical terms, the equation is expressed as:

$$C^2 = A^2 + B^2 \tag{5.19}$$

where A and B are the lengths of two line segments that are not the hypotenuse (C). Therefore, if you know the projected X, Y coordinates for any two points, it is a relatively straightforward matter to determine the distance between them—you simply determine the lengths of the two legs and then calculate the length of the hypotenuse. For example, suppose you want to measure the distance between Point #1 and Point #2 in the Cartesian coordinate system shown in Figure 5-16a and listed in Table 5-5. To determine the distance between these two points it is first necessary to calculate the distance of the two sides of the triangle that form the right angle. To determine the length of line A, one of the Y values is subtracted from the other Y value (2 – 5 = –3). To determine length of line B, one of the X values is subtracted from the other X value (2 – 7 = –5). Both values are used in Equation 5.19:

TABLE 5–5 The coordinates of Point #1 and Point #2 in Figure 5-16 in meters.

Point	X-coordinate	Y-coordinate
1	2	2
2	7	5

$$C^2 = (-3)^2 + (-5)^2$$

$$C^2 = 9 + 25$$

$$C^2 = 34$$

$$C = \sqrt{34}$$

$$C = 5.83 \text{ m}$$

The distance between Point #1 and Point #2 in Figure 5-16a is 5.83 m.

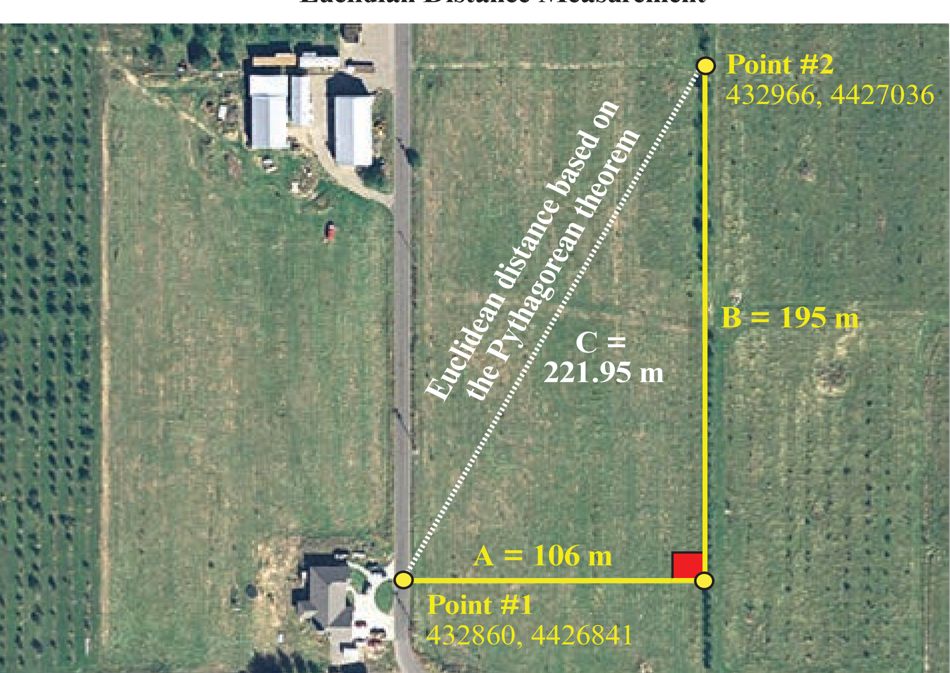


FIGURE 5–17 Euclidean distance measurements are based on the Pythagorean theorem where the lengths of lines A and B in a right triangle can be used to calculate the length of the hypotenuse, C. In this example, the distance from a point in front of the single-family residence to a corner of the adjacent field is calculated (aerial photography courtesy state of Utah).

TABLE 5–6 The coordinates of Point #1 and Point #2 in Figure 5-17 in a Universal Transverse Mercator (UTM) Cartesian coordinate system (Zone 12 N).

Point	UTM *X*-coordinate	UTM *Y*-coordinate
1	432,860	4,426,841
2	432,966	4,427,036

Normally, the coordinates of the points that you are investigating are described in terms of their location in a particular coordinate system, e.g., the Universal Transverse Mercator (UTM) map projection. For example, suppose you wanted to compute the distance between Point #1 and Point #2 in the agricultural field shown in Figure 5-17. The coordinates of the two points are listed in Table 5-6.

In this case the computation would be:

$$C^2 = (432860 - 432966)^2 + (4426841 - 4427036)^2$$

$$C^2 = (-106)^2 + (-195)^2$$

$$C^2 = 11236 + 38025$$

$$C^2 = 49261$$

$$C = \sqrt{49261}$$

$$C = 221.95 \text{ m}$$

A limitation of using the Pythagorean theorem is that this method only works when the points used in the measurement are comparable—that is, they are in a coordinate system based on a projection such as UTM. Also, this method will not work for latitude and longitude or for projected points that are great distances from one another because of the curvature of the Earth.

Manhattan Distance Measurement

Linear distance measurement based on the Pythagorean theorem is useful for many applications, but it does have limitations. One limitation is the fact that to get from Point #1 to Point #2 in an urban environment—and even in many natural settings—it may not be logical to measure the Euclidean distance between two points using a straight line (i.e., using the hypotenuse) "as the crow flies." To remedy this, the Manhattan distance between two points may be calculated, where:

$$\text{Manhattan}_{\text{distance}} = |X_1 - X_2| + |Y_1 - Y_2| \quad (5.20)$$

Euclidean Distance versus Manhattan Distance Measurement

FIGURE 5–18 The white line from the intersection at location A to the intersection at location B would require you to walk through commercial buildings, climb over fences, and pass through parks and natural areas. Therefore, the Euclidean distance (white line) would not be representative of the practical distance to walk from location A to location B. Conversely, it is possible to determine the Manhattan "round-the-block" distance for the same journey. In this example, all of the lines except for the white line, have exactly the same length from location A to location B as they traverse through the city (aerial photography courtesy of state of Utah).

The **Manhattan distance** (sometimes referred to as "round-the-block" or "city block" distance) between two points utilizes the lengths of the two sides of the right triangle, but not the hypotenuse (Figure 5-16b). This is analogous to going from Point #1 to Point #2 in a city where you cannot simply walk through buildings or climb over fences. Rather, you are required to walk around the block to get from Point #1 to Point #2. For example, the Manhattan distance between Point #1 and Point #2 in Figure 5-16b is:

$$\text{Manhattan}_{\text{distance}} = |2 - 7| + |2 - 5|$$

$$\text{Manhattan}_{\text{distance}} = 5 + 3$$

$$\text{Manhattan}_{\text{distance}} = 8 \text{ m}$$

Note that this is *not* the same value (5.83 m) calculated using the Pythagorean theorem in the previous example (Figure 5-16a).

Therefore, while it is often useful to determine the Euclidean distance between Point #1 and Point #2 "as the crow flies" using the Pythagorean theorem, other factors must sometimes be considered before a practical distance measurement can be made. In fact, many geographic areas have features that *impede* making direct Pythagorean theorem-based distance measurements between two points.

For example, suppose we need to go from location A to location B in downtown Salt Lake City, UT (Figure 5-18). We are interested in traveling the shortest distance

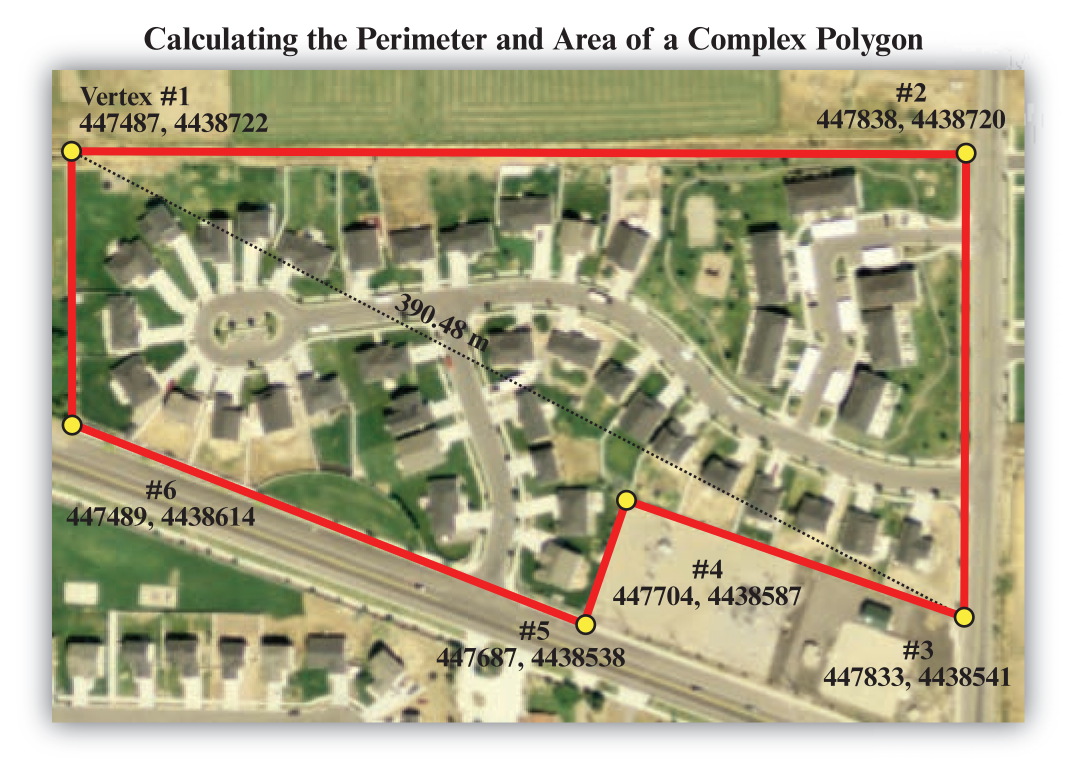
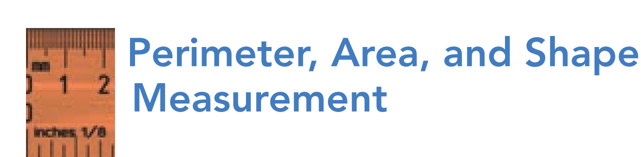


FIGURE 5–19 A six-sided polygon circumscribes a subdivision located in Spanish Fork, UT (vertex #1 is used twice to close the polygon). The perimeter and area of this polygon can be calculated using Equations 5.21 and 5.22, respectively. The length of the longest diagonal that can be inscribed in the polygon is 390.48 m based on the Pythagorean theorem. The area and longest diagonal information can be used to compute the compactness of the polygon (aerial photography courtesy state of Utah).

possible between location A and B. The white line is the hypotenuse of a right triangle and is the most direct and shortest route between the two locations. In this case, however, the Euclidean distance between location A and B is actually of little value because it would be necessary to walk through buildings, climb over fences, and walk through numerous stands of vegetation to get from one location to another, even though it is the shortest distance.

A more suitable approach to obtaining an accurate measurement of the functional or practical distance between location A and B is to a) identify a route that takes into account existing roads and trails and subdivide it into n (e.g., 6) logical line segments, and then b) compute the distance of these n segments using either Pythagorean theorem or Manhattan distance logic. The more practical distance between location A and location B would be one of the colored lines drawn in Figure 5-18.

Numerous digital image processing programs allow the analyst to specify whether distances are to be computed in multispectral feature space using a) the Pythagorean theorem, or b) the Manhattan distance. For example, the minimum distance to mean classification discussed in Chapter 8 can be based on either type of distance measurement.



Perimeter, Area, and Shape Measurement

Perimeter Measurement

The **perimeter** of a polygon is measured by determining the length of each of the n line segments associated with a polygon and then summing them:

$$Perimeter = \sum_{i=1}^{n} length_i \tag{5.21}$$

The length of each line segment is usually computed using the Pythagorean theorem as previously discussed. For example, consider the six-sided polygon with six vertices circumscribing a subdivision in Spanish Fork, UT, shown in Figure 5-19. The coordinates of the six polygon vertices are listed in Table 5-7. To determine the perimeter of this polygon, the Pythagorean theorem was used to compute the distance associated

TABLE 5–7 UTM coordinates of six vertices surrounding a subdivision in Spanish Fork, UT, (Figure 5-19) used to demonstrate how to calculate the perimeter of the complex polygon.

Vertex	UTM *X*-coordinate Easting	UTM *Y*-coordinate Northing	Distance from Vertex to Vertex (e.g., 1 to 2)
1	447487	4438722	
2	447838	4438720	351.01
3	447833	4438541	179.07
4	447704	4438587	136.96
5	447687	4438538	51.87
6	447489	4438614	212.08
1	447487	4438722	108.02
	$Perimeter = \sum_{i=1}^{n} length_i =$		**1,039 m**

with each line segment. The perimeter of the polygon is 1,039 m.

Area Measurement

The area of a rectangle on a remotely sensed image is computed simply by multiplying the values of its length and width, i.e., $A = l \times w$. Another elementary area computation is that of a circle, which is $A = \pi r^2$. Complications can arise, however, when the shape of the polygon varies from a rectangle or circle. In the remote sensing literature, polygons are also often referred to as **areas of interest** (**AOIs**).

To calculate the area of a polygon (or AOI) in remotely sensed imagery, the analyst typically uses a rubber-band tool to identify n vertices at unique map (x,y) or image (row and column) coordinates. The "contribution" of each point (vertex) in the polygon to the area is computed by evaluating the x-coordinate of a vertex prior to the vertex under examination (x_{i-1}) with the x-coordinate of the next vertex in the sequence (x_{i+1}) and multiplying the result by the y-coordinate (y_i) of the vertex under examination according to the formula:

$$Area = 0.5\left|\sum_{i=1}^{n} y_i(x_{i+1} - x_{i-1})\right| . \quad (5.22)$$

It is important to point out that the very first vertex in the polygon uses values from the second and the nth (last) vertex. The last vertex uses values from the $n-1$ vertex and the first vertex.

For example, consider the use of Equation 5.22 to compute the area of the complex polygon surrounding the residential subdivision in Spanish Fork, UT (Figure 5-19). Table 5-8 presents the X- and Y-coordinates of the six vertices and the contribution calculated for each vertex. The area of the polygon is 52,216 m^2 or approximately 5.2 hectares. This example only has six vertices. As you might imagine, it is more computationally intensive to calculate the area of an extremely complex polygon that has hundreds or even thousands of vertices. Also, some polygons may have polygons inside them. In such cases, the area of the enclosed polygon needs to be calculated and then subtracted from the area of the surrounding polygon.

After calculating the basic perimeter and area properties of various polygons (patches) throughout a landscape, it is possible to use these properties to compute numerous landscape ecology metrics often used in geographic research (e.g., Frohn and Hao, 2006).

Polygon (AOI) area estimates extracted from rectified remote sensor data are more accurate than area estimates extracted from unrectified imagery. Unrectified

TABLE 5–8 UTM coordinates of the six vertices surrounding a subdivision in Spanish Fork, UT, (Figure 5-19) used to demonstrate how to calculate the area of a complex polygon.

Vertex	UTM X-coordinate Easting	UTM Y-coordinate Northing	Contribution
6	447489	4438614	
1	447487	4438722	1549113978
2	447838	4438720	1535797120
3	447833	4438541	–594764494
4	447704	4438587	–648033702
5	447687	4438538	–954285670
6	447489	4438614	–887722800
1	447487	4438722	
	$Area = 0.5\left\lvert\sum_{i=1}^{n} y_i(x_{i+1} - x_{i-1})\right\rvert =$		**52,216 m^2**

imagery contains geometric error that can severely affect the accuracy of the areal estimates.

Shape Measurement

Polygonal areas of interest all have two-dimensional *shapes* with constant relationships of position and distance of the vertices on their perimeters. One way to measure the shape of an AOI is to compute an index that relates the real-world shape (e.g., the shape of the six-sided polygon shown in Figure 5-19) to some regular geometric figure such as a circle or hexagon. The most commonly used shape index is *compactness,* which is based on deviations from the most compact shape, a circle. The compactness shape index assumes a value of 1.0 for a perfect circle (i.e., a circle has maximum compactness) and a value <1.0 for any less compact shape. At the opposite end of the range of the shape index is the straight line, which has no area and a shape index of zero (Earickson and Harlin, 1994).

The shape (S) of an area of interest may be computed using the equation

$$S = \frac{2\sqrt{(A \div \pi)}}{l}, \qquad (5.23)$$

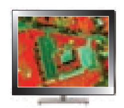
where l is the length of the longest diagonal of the shape and A is the area. In this example, the longest diagonal within the polygon is from vertex #1 to vertex #3 which equals 390.48 m. The area of the polygon is 52,216 m^2.

The shape of the six-sided polygon when compared to a perfect circle is

$$S = \frac{2\sqrt{(52216 \div 3.1416)}}{390.48}$$

$$S = \frac{257.84}{390.48} = 0.66 .$$

The shape index value of the residential subdivision is moderately compact.



References

Aiazzi, B., Baronti, S. and M. Selva, 2007, "Improving Component Substitution Pansharpening Through Multivariate Regression of MS+Pan Data," *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3230–3239.

Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P., and L. M. Bruce, 2007, "Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data-Fusion Contest," *IEEE Transactions on Geoscience and Remote Sensing*, 45(10): 3012–3021.

Beauchemin, M., and K. B. Fung, 2001, "On Statistical Band Selection for Image Visualization," *Photogrammetric Engineering & Remote Sensing*, 67(5):571–574.

Campbell, J., 2007, *Introduction to Remote Sensing*, NY: Guilford Press, 626 p.

Carper, W. J., Kiefer, R. W., and T. M. Lillesand, 1990, "The Use of Intensity–Hue–Saturation Transformation for Merging SPOT Panchromatic and Multispectral Image Data," *Photogrammetric Engineering & Remote Sensing*, 56(4):459–467.

Chavez, P. S., 1986, "Digital Merging of Landsat TM and Digitized NHAP Data for 1:24,000 Scale Image Mapping," *Photogrammetric Engineering & Remote Sensing*, 56(2):175–180.

Chavez, P. S., and J. A. Bowell, 1988, "Comparison of the Spectral Information Content of Landsat Thematic Mapper and SPOT for Three Different Sites in the Phoenix, Arizona, Region," *Photogrammetric Engineering & Remote Sensing*, 54(12):1699–1708.

Chavez, P. L., Berlin, G. L., and L. B. Sowers, 1982, "Statistical Method for Selecting Landsat MSS Ratios," *Journal of Applied Photographic Engineering*, 8(1):23–30.

Chavez, P. S., Guptill, S. C., and J. A. Bowell, 1984, "Image Processing Techniques for Thematic Mapper Data," *Proceedings—ASPRS Technical Papers*, 2:728–742.

Chavez, P. S., Sides, S. C., and J. A. Anderson, 1991, "Comparison of Three Different Methods to Merge Multiresolution and Multispectral Data: Landsat TM and SPOT Panchromatic," *Photogrammetric Engineering & Remote Sensing*, 57(3):295–303.

Chen, C. M., Hepner, G. F., and R. R. Forster, 2003, "Fusion of Hyperspectral and Radar Data Using the IHS Transformation to Enhance Urban Surface Features," *ISPRS Journal of Photogrammetry & Remote Sensing*, 58:19–30.

Chen, F., Guan, Z., Yang, X., and W. Cui, 2011, "A Novel Remote Sensing Image Fusion Method based on Independent Component Analysis," *International Journal of Remote Sensing*, 32(10):2745–2763.

Choi, M., 2006, "A New Intensity-Hue-Saturation Fusion Approach to Image Fusion with a Trade-off Parameter," *IEEE Transactions on Geoscience and Remote Sensing*, 44(6):1672–1682.

Earickson, R., and J. Harlin, 1994, *Geographic Measurement and Quantitative Analysis*, New York: Macmillan, 350 p.

Earnshaw, R. A., and N. Wiseman, 1992, *An Introduction Guide to Scientific Visualization,* NY: Springer, 156 p.

Ehlers, M., Jadkowski, M. A., Howard, R. R., and D. E. Brostuen, 1990, "Application of SPOT Data for Regional Growth Analysis and Local Planning," *Photogrammetric Engineering & Remote Sensing*, 56(2):175–180.

Frohn, R. C., and Y. Hao, 2006, "Landscape Metric Performance in Analyzing Two Decades of Deforestation in the Amazon Basin of Rondonia, Brazil," *Remote Sensing of Environment*, 100:237–251.

Gonzalez-Audicana, M., Salete, J. L., Catalan, R. G., and R. Garcia, 2004, "Fusion of Multispectral and Panchromatic Images Using Improved IHS and PCA Mergers Based on Wavelet Decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, 42(6):1291–1299.

Jalan, S., and B. S. Sokhi, 2012, "Comparison of Different Pan-sharpening Methods for Spectral Characteristic Preservation: Multi-temporal CARTOSAT-1 and IRS-P6 LISS-IV Imagery," *International Journal of Remote Sensing*, 33(18):5629–5643.

Jensen, J. R., and R. R. Jensen, 2013, *Introductory Geographic Information Systems*, Boston: Pearson, 400 p.

Khaleghi, B., Khamis, A., Karry, F.O., and S. N. Razavi, 2013, "Multisensor Data Fusion: A Review of the State of the Art," *Information Fusion*, 14(1):28–44.

Klonus, S., and M. Ehlers, 2007, "Image Fusion using the Ehlers Spectral Characteristics Preservation Algorithm," *GIScience & Remote Sensing*, 44:93–116.

Klonus, S., and M. Ehlers, 2009, "Performance of Evaluation Methods in Image Fusion," *Proceedings of the 12th International Conference on Information Fusion*, Seattle, July 6–9, 2009, 8 p.

Konecny, G., 2014, *Geoinformation: Remote Sensing, Photogrammetry and Geographic Information Systems*, Boca Raton: CRC Press, 416 p.

Kulkarni, A. D., 2001, *Computer Vision and Fuzzy-Neural Systems*, Upper Saddle River: Prentice-Hall, 504 p.

Laban, C. A., Bernard, V., and W. Brower, 2000, *Process for Enhancing the Spatial Resolution of Multispectral Imagery using Pan-sharpening,* US Patent 6,011,875.

Lillesand, T. M., Kiefer, R. W., and J. W. Chipman, 2008, *Remote Sensing and Image Interpretation*, 6th Ed., NY: John Wiley, 756 p.

Liu, J. G., 2000a, "Evaluation of Landsat-7 ETM^+ Panchromatic Band for Image Fusion with Multispectral Bands," *Natural Resources Research*, 9(4):269–276.

Liu, J. G., 2000b, "Smoothing Filter Based Intensity Modulation: A Spectral Preserving Image Fusion Technique for Improving Spatial Details," *International Journal of Remote Sensing*, 21(18):3461–3472.

Liu, Z., Blasch, E., Xue, Z., Zhao, J., Laganiere, R., and W. Wu, 2012, "Objective Assessment of Multiresolution Image Fusion Algorithms for Context Enhancement in Night Vision: A Comparative Study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1):94–109.

Meng, Q., Borders, B., and M. Madden, 2010, "High-resolution Satellite Image Fusion using Regression Kriging," *International Journal of Remote Sensing*, 31(7):1857–1876.

Myler, H. R., 2013, "Module 1.9: Photonic Devices for Imaging, Display, and Storage," in *Fundamentals of Photonics*, SPIE International Society for Optics and Photonics, 349–380, http://spie.org/x17229.xml.

Park, J. H., and M. G. Kang, 2004, "Spatially Adaptive Multi-resolution Multispectral Image Fusion, *International Journal of Remote Sensing*, 25(23):5491–5508.

Pellemans, A. H., Jordans, R. W., and R. Allewijn, 1993, "Merging Multispectral and Panchromatic SPOT Images with Respect to the Radiometric Properties of the Sensor," *Photogrammetric Engineering & Remote Sensing*, 59(1):81–87.

Pratt, W. K., 2013, *Introduction to Digital Image Processing*, Boca Raton: CRC Press, 736 p.

Prost, G. L., 2013, *Remote Sensing for Geoscientists: Image Analysis and Integration*, 3rd Ed., Boca Raton: CRC Press, 702 p.

Ready, J., 2013, "Module 1.6: Optical Detectors and Human Vision," in *Fundamentals of Photonics*, SPIE International Society for Optics and Photonics, 349–380, http://spie.org/x17229.xml.

Richards, J. A., 2013, *Remote Sensing Digital Image Analysis: An Introduction*, 5th Ed., NY: Springer-Verlag, 494 p.

Sabins, F. F., 2007, *Remote Sensing: Principles and Interpretation*, 3rd Ed., San Francisco: W. H. Freeman.

Schowengerdt, R. A., 1980, "Reconstruction of Multispatial, Multispectral Image using Spatial Frequency Contents," *Photogrammetric Engineering & Remote Sensing*, 46:1325–1334.

Schowengerdt, R. A., 2006, *Remote Sensing: Models and Methods for Image Processing,* 3rd Ed., NY: Academic Press, 515 p.

Shah, V. P., Younan, N. H., and R. L. King, 2008, "An Efficient Pan-Sharpening Method via a Combined Adaptive PCA Approach and Contourlets," *IEEE Transactions on Geoscience and Remote Sensing*, 46(5):1232–1335.

Sheffield, C., 1985, "Selecting Band Combinations from Multispectral Data," *Photogrammetric Engineering & Remote Sensing*, 51(6):681–687.

Slocum, T. A., McMaster, R.B., Kessler, F. C., and H. H. Howard, 2008, *Thematic Cartography and Geographic Visualization*, 3rd Ed., Upper Saddle River: Prentice Hall.

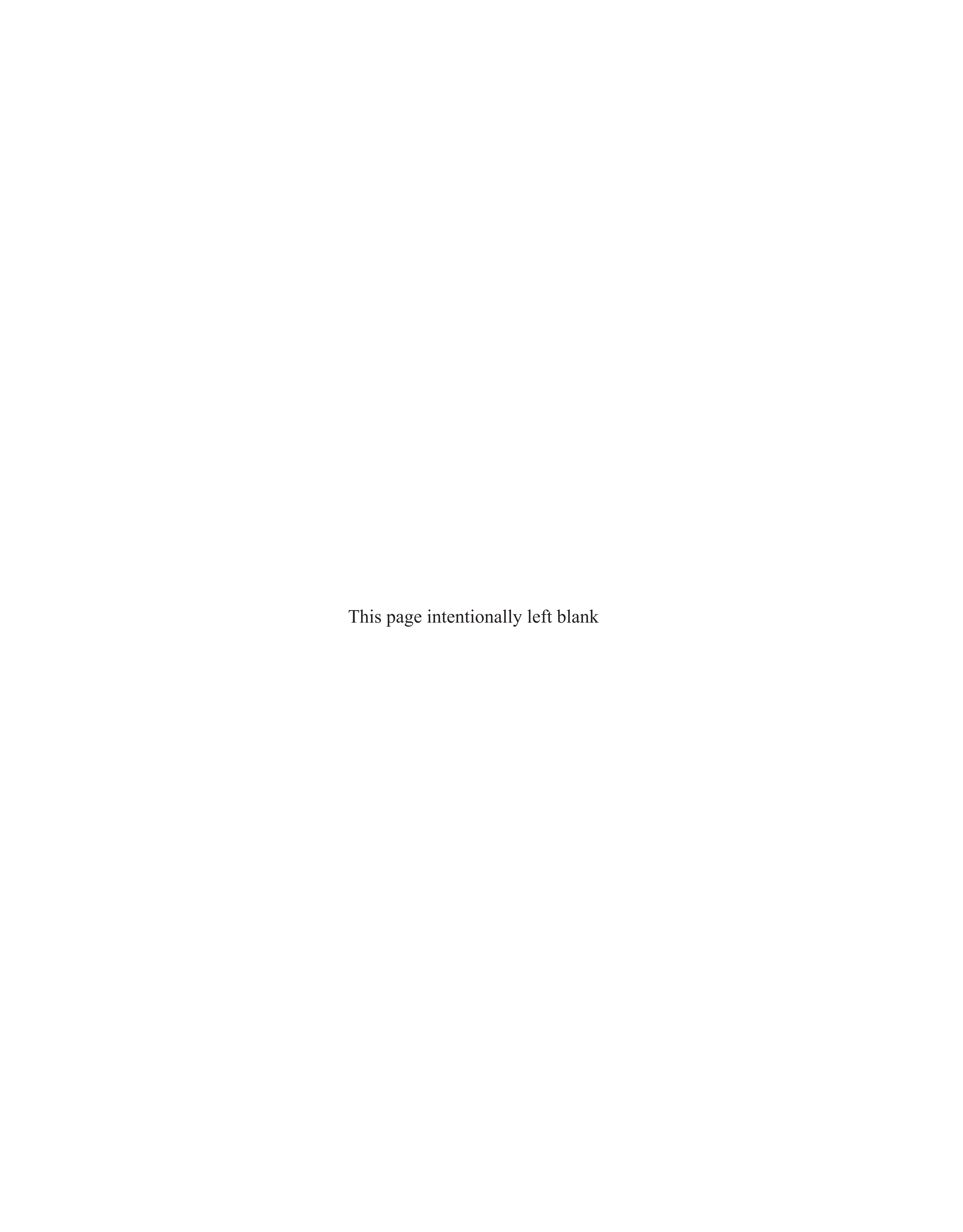
SPIE, 2013, *Fundamentals of Photonics*, SPIE International Society for Optics & Photonics, http://spie.org/x17229.xml.

Van der Meer, F., 2012, "Remote Sensing Image Analysis and Geostatistics," *International Journal of Remote Sensing*, 33(18):5644–5676.

Yang, W., and Y. Gong, 2012, Multi-spectral and Panchromatic Images Fusion based on PCA and Fractional Spline Wavelet," *International Journal of Remote Sensing*, 33(20):7060–7074.

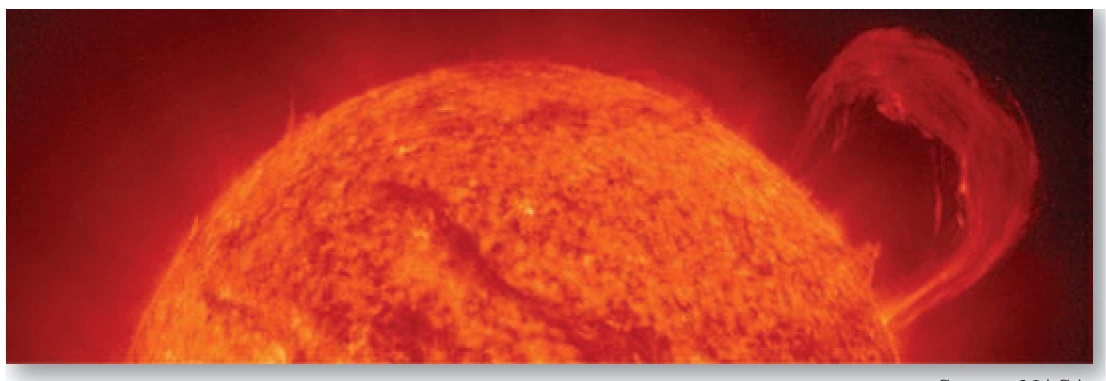
Yao, W., and M. Han, 2010, "Improved GIHSA for Image Fusion based on Parameter Optimization," *International Journal of Remote Sensin*g, 31(10):2717–2728.

Zhu, Y., Varshney, P. K., and H. Chen, 2011, "ICA-based Fusion for Colour Display of Hyperspectral Images," *International Journal of Remote Sensing*, 32(9):2427–2450.



This page intentionally left blank

6 Electromagnetic Radiation Principles and Radiometric Correction



Source: NASA

The perfect remote sensing system has yet to be developed. Also, the Earth's atmosphere, land, and water are very complex and do not lend themselves well to being recorded by remote sensing instruments that have constraints such as spatial, spectral, temporal, and radiometric resolution. Consequently, error and noise can be introduced into the data acquisition process that can degrade the quality of the remote sensor data collected (Congalton and Green, 2009). This in turn may have an impact on the accuracy of subsequent human- or machine-assisted image analysis (Song et al., 2001).

The two most common types of error encountered in remotely sensed data are radiometric and geometric error. **Radiometric correction** is concerned with improving the accuracy of surface spectral reflectance, emittance, or back-scattered measurements obtained using a remote sensing system (Johannsen and Daughtry, 2009; San and Suzen, 2010). **Geometric correction** is concerned with placing the reflected, emitted, or back-scattered measurements or derivative products in their proper planimetric (map) location. The data can then be associated with other geospatial information in a geographic information system (GIS) or in a spatial decision support system (SDSS) (Jensen and Jensen, 2013).

Radiometric and geometric correction of remotely sensed data are normally referred to as **preprocessing** operations because they are performed prior to information extraction. Image preprocessing ideally produces a corrected image that is as close as possible, both radiometrically and geometrically, to the true radiant energy and spatial characteristics of the study area at the time of data collection. Internal and external errors must be identified to correct the remotely sensed data:

- *Internal errors* are introduced by the remote sensing system itself. They are generally systematic (predictable) and may be identified and then corrected based on prelaunch or in-flight calibration measurements. For example, *n*-line striping in the imagery may be caused by a single detector that has become uncalibrated. In many instances, radiometric correction can adjust for detector miscalibration.
- *External errors* are usually introduced by phenomena that vary in nature through space and time. The most important external variables that can cause remote sensor data to exhibit radiometric and geometric error are the atmosphere, terrain elevation, slope, and aspect. Some external errors may be corrected by relating empirical observations made on the ground (i.e., radiometric and geometric ground control points) to sensor system measurements.



Overview

This chapter focuses on *radiometric* correction of remotely sensed data. Radiometric correction requires knowledge about electromagnetic radiation principles and what interactions take place during the remote sensing data collection process. To be exact, it also involves knowledge about the terrain slope and aspect and bidirectional reflectance characteristics of the scene. Therefore, this chapter first reviews fundamental electromagnetic radiation principles. It then discusses how these principles and relationships are used to correct for radiometric distortion in remotely sensed data

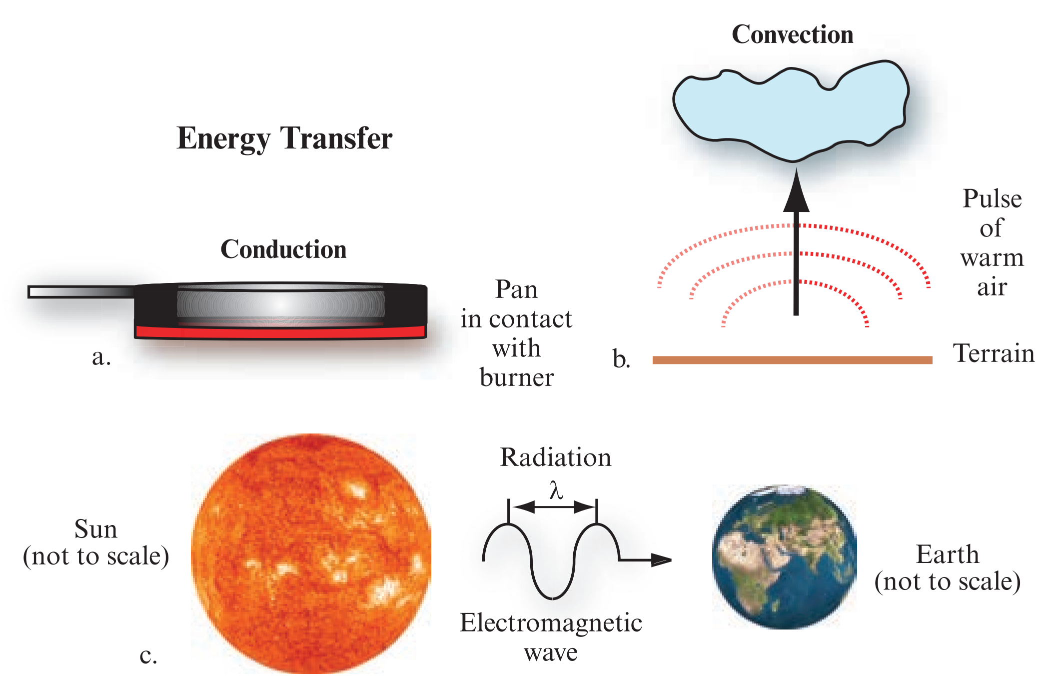
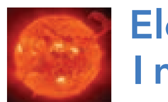


FIGURE 6–1 Energy may be transferred three ways: conduction, convection, and radiation. a) Energy may be conducted directly from one object to another as when a pan is in direct physical contact with a hot burner. b) The Sun bathes the Earth's surface with radiant energy causing the air near the ground to increase in temperature. The less dense air rises, creating convectional currents in the atmosphere. c) Electromagnetic energy in the form of electromagnetic waves may be transmitted through the vacuum of space from the Sun to the Earth.

caused primarily by the atmosphere and elevation. Chapter 7 discusses the geometric correction of remote sensor data.



Electromagnetic Energy Interactions

Energy recorded by remote sensing systems undergoes fundamental interactions that should be understood to properly preprocess and interpret remotely sensed data. For example, if the energy being remotely sensed comes from the Sun, the energy:

- is radiated by atomic particles at the source (the Sun),
- travels through the vacuum of space at the speed of light,
- interacts with the Earth's atmosphere,
- interacts with the Earth's surface,
- interacts with the Earth's atmosphere once again, and
- finally reaches the remote sensing system, where it interacts with various optics, filters, film emulsions, or detectors.

Each of these interactions has an impact on the radiometric quality of the information stored in the remote sensor data. It is instructive to examine each of these interactions that electromagnetic energy undergoes as it progresses from its source to the remote sensing system detector.



Conduction, Convection, and Radiation

Energy is the ability to do work. In the process of doing work, energy may be transferred from one body to another or from one place to another. The three basic ways in which energy can be transferred are conduction, convection, and radiation (Figure 6-1). Most people are familiar with *conduction,* which occurs when one body (molecule or atom) transfers its kinetic energy to another by colliding with it. This is how a metal pan is heated by a hot burner on a stove. During *convection*, the kinetic energy of bodies is transferred from one place to another by physically moving the bodies. A good example is the heating of the air near the ground in the morning hours. The warmer air near the surface rises, setting up convectional currents in the atmosphere, which may produce cumulus clouds. The transfer of energy by electromagnetic *radiation* is of primary interest to remote sensing science because it is the only form of energy transfer that can take place in a vacuum such as the region between the Sun and the Earth.

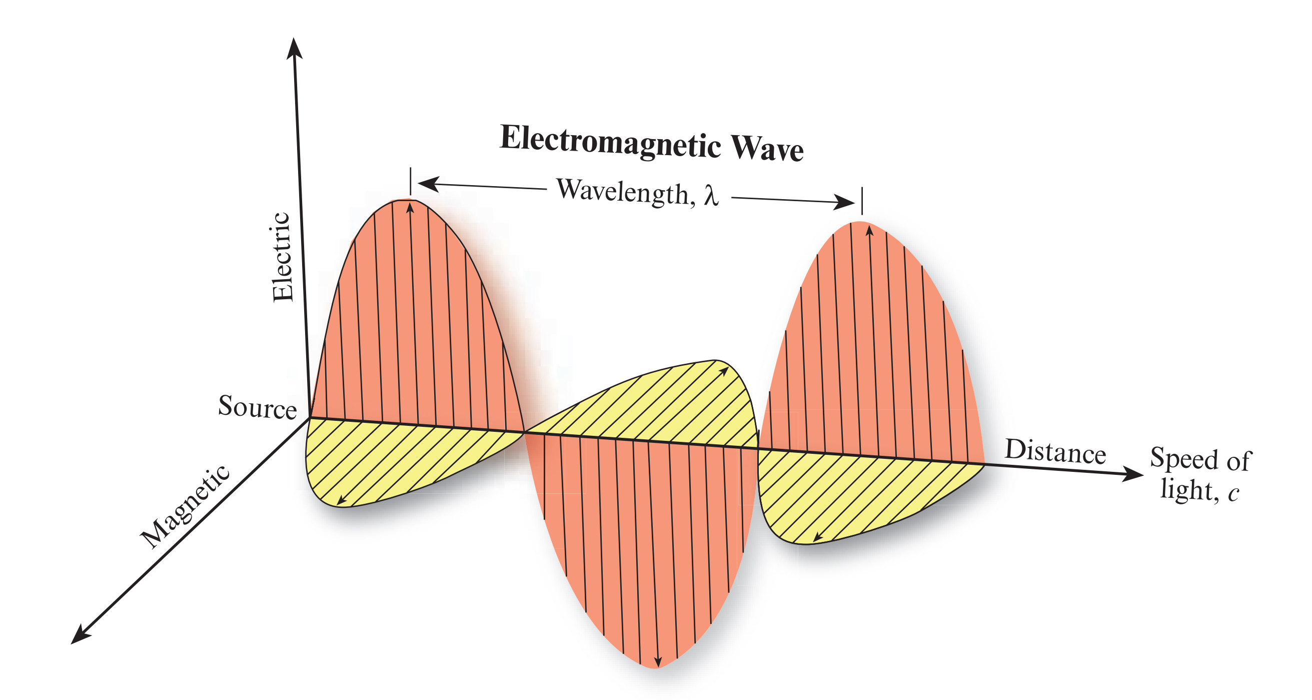
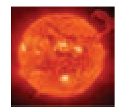


FIGURE 6-2 An electromagnetic wave is complex and composed of both electric and magnetic vectors that are orthogonal (at 90° angles) to one another. The waves travel from the source at the speed of light (3×10^8 m/s).



Electromagnetic Radiation Models

To understand how electromagnetic radiation is created, how it propagates through space, and how it interacts with other matter, it is useful to describe the processes using two different models: the *wave* model and the *particle* model (Vandergriff, 2014).

Wave Model of Electromagnetic Energy

In the 1860s, James Clerk Maxwell (1831–1879) conceptualized electromagnetic radiation (EMR) as an electromagnetic wave that travels through space at the speed of light, *c*, which is 2.99×10^8 meters per second (hereafter referred to as m/s) or 186,282.03 miles per second. A useful relation for quick calculations is that light travels about 1 ft per nanosecond (10^{-9} s). Light travels in a vacuum at a constant speed, and this speed is considered a universal constant. It is important to note that speed changes for light traveling through non-vacuum media such as air (0.03% slower) or glass (30.0% slower) (Vandergriff, 2014). The *electromagnetic wave* consists of two fluctuating fields—one electric and the other magnetic (Figure 6-2). The two vectors are at right angles (i.e., orthogonal) to one another, and both vectors are perpendicular to the direction of travel.

How is an electromagnetic wave created? *Electromagnetic radiation* is generated whenever an electrical charge is accelerated. The wavelength (λ) of the electromagnetic radiation depends upon the length of time that the charged particle is accelerated. Its frequency (ν) depends on the number of accelerations per second. *Wavelength* is formally defined as the mean distance between maximums (or minimums) of a roughly periodic pattern (Figures 6-2 and 6-3) and is normally measured in micrometers (μm) or nanometers (nm). *Frequency* is the number of wavelengths that pass a point per unit time. A wave that sends one crest by every second (completing one cycle) is said to have a frequency of one cycle per second, or one *hertz*, abbreviated 1 Hz. Frequently used measures of wavelength and frequency are found in Table 6-1.

The relationship between the wavelength (λ) and frequency (ν) of electromagnetic radiation is based on the following formula, where *c* is the speed of light:

$$c = \lambda\nu, \tag{6.1}$$

$$\nu = \frac{c}{\lambda}, \tag{6.2}$$

and

$$\lambda = \frac{c}{\nu}. \tag{6.3}$$

Note that frequency is *inversely* proportional to wavelength. This relationship is shown diagrammatically in

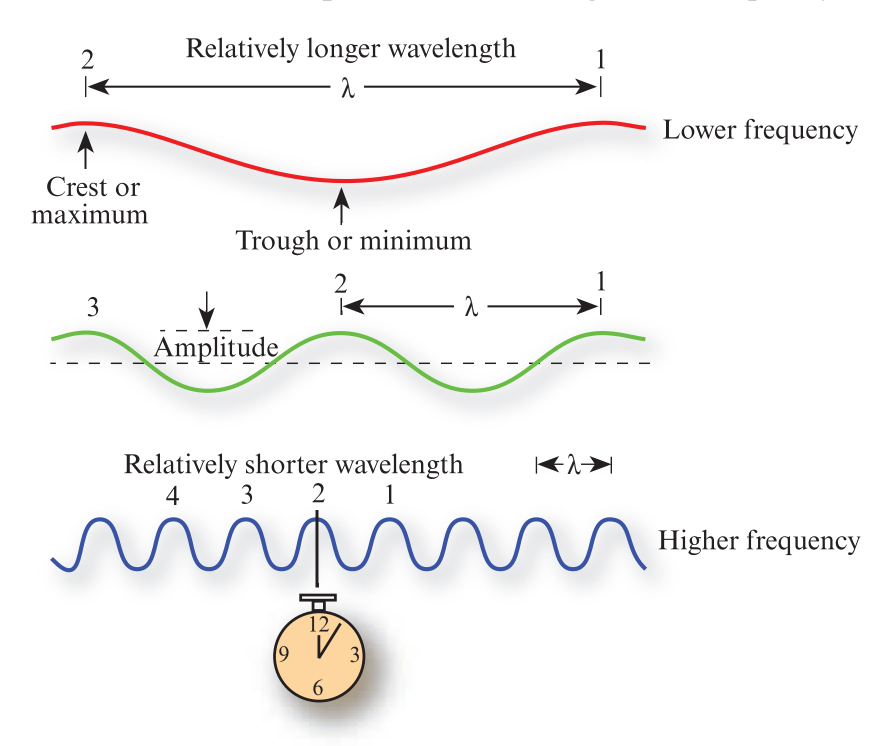


FIGURE 6–3 This cross-section of several electromagnetic waves illustrates the inverse relationship between wavelength (λ) and frequency (ν). The longer the wavelength, the lower the frequency; the shorter the wavelength, the higher the frequency. The amplitude of an electromagnetic wave is the height of the wave crest above the undisturbed position. Successive wave crests are numbered 1, 2, 3, and 4. An observer at the position of the clock records the number of crests that pass by in a second. This frequency is measured in cycles per second or *hertz*.

Figure 6-3, where the longer the wavelength, the lower the frequency; the shorter the wavelength, the higher the frequency (Seeber, 2014). When electromagnetic radiation passes from one substance to another, the speed of light and wavelength change while the frequency remains the same.

All objects above absolute zero (–273°C or 0 K) emit electromagnetic energy, including water, soil, rock, vegetation, and the surface of the Sun. The Sun represents the initial source of most of the electromagnetic energy recorded by remote sensing systems (except RADAR, LiDAR, and SONAR) (Figures 6-4 and 6-5). We may think of the Sun as a 5,770 – 6,000 K *blackbody* (a theoretical construct that absorbs and radiates energy at the maximum possible rate per unit area at each wavelength (λ) for a given temperature). The total emitted radiation from a blackbody (M_λ) measured in watts per m^{-2} is proportional to the fourth power of its absolute temperature (T) measured in kelvin (K). This is known as the *Stefan-Boltzmann law* and is expressed as:

$$M_\lambda = \sigma T^4 \tag{6.4}$$

where σ is the Stefan-Boltzmann constant which equals $5.6697 \times 10^{-8} \mathrm{W\ m^{-2} K^{-4}}$. The important thing to remember is that the amount of energy emitted by an object such as the Sun or the Earth is a function of its temperature. The greater the temperature, the greater the amount of radiant energy exiting the object. The actual amount of energy emitted by an object is computed by summing (integrating) the area under its curve (Figure 6-6). It is clear from this illustration that the total emitted radiation from the 6,000 K Sun is far greater than that emitted by the 300 K Earth.

In addition to computing the total amount of energy exiting a theoretical blackbody such as the Sun, we can determine its dominant wavelength (λ_{max}) based on *Wien's displacement law*:

$$\lambda_{max} = \frac{k}{T} \tag{6.5}$$

where k is a constant equaling 2,898 μm K, and T is the absolute temperature in kelvin. Therefore, as the Sun approximates a 6,000 K blackbody, its dominant wavelength (λ_{max}) is 0.48 μm:

TABLE 6–1 Wavelength and frequency units of measurement.

Wavelength (λ)		Frequency (cycles per second)	
kilometer (km)	1,000 m	hertz (Hz)	1
meter (m)	1.0 m	kilohertz (kHz)	$1{,}000 = 10^3$
centimeter (cm)	$0.01\ m = 10^{-2}$ m	megahertz (MHz)	$1{,}000{,}000 = 10^6$
millimeter (mm)	$0.001\ m = 10^{-3}$ m	gigahertz (GHz)	$1{,}000{,}000{,}000 = 10^9$
micrometer (μm)	$0.000001 = 10^{-6}$ m		
nanometer (nm)	$0.000000001 = 10^{-9}$ m		
angstrom (A)	$0.0000000001 = 10^{-10}$ m		

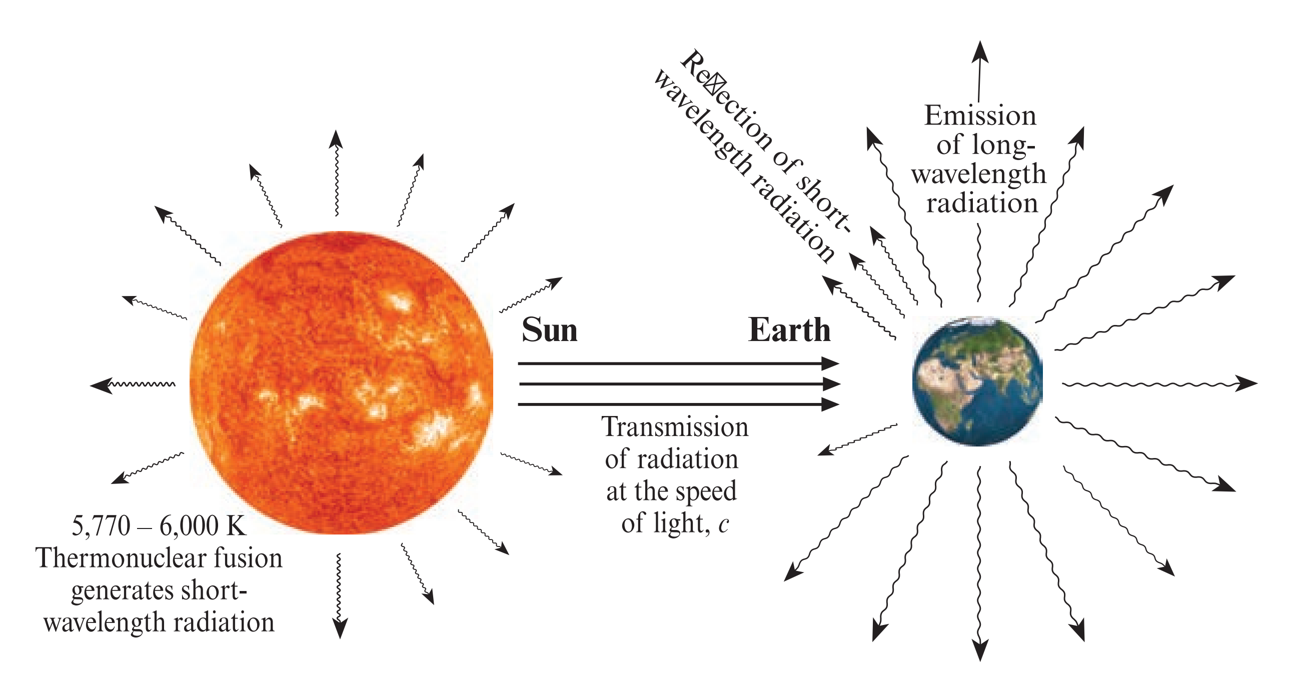


FIGURE 6–4 The thermonuclear fusion taking place on the surface of the Sun yields a continuous spectrum of electromagnetic energy. The 5,770 – 6,000 kelvin (K) temperature of this process produces a large amount of relatively short wavelength energy that travels through the vacuum of space at the speed of light. Some of this energy is intercepted by the Earth, where it interacts with the atmosphere and surface materials. The Earth reflects some of the energy directly back out to space or it may absorb the short wavelength energy and then re-emit it at a longer wavelength (images courtesy of NASA; line art based on Strahler and Strahler, 1989).

$$0.483\,\mu m = \frac{2{,}898\ \mu m\,K}{6{,}000\ K}.$$

Electromagnetic energy from the Sun travels 8 minutes across the intervening 93 million miles (150 million km) of space to reach the Earth. As shown diagrammatically in Figure 6-6, the Earth approximates a 300 K (27°C) blackbody and has a dominant wavelength at approximately 9.66 μm:

$$9.66\ \mu m = \frac{2{,}898\ \mu m\,K}{300\ K}.$$

Although the Sun has a dominant wavelength at 0.48 μm, it produces a continuous spectrum of electromagnetic radiation ranging from very short, extremely high-frequency gamma and cosmic waves to long, very-low-frequency radio waves (Figures 6-7 and 6-8). The Earth intercepts only a very small portion of the electromagnetic energy produced by the Sun.

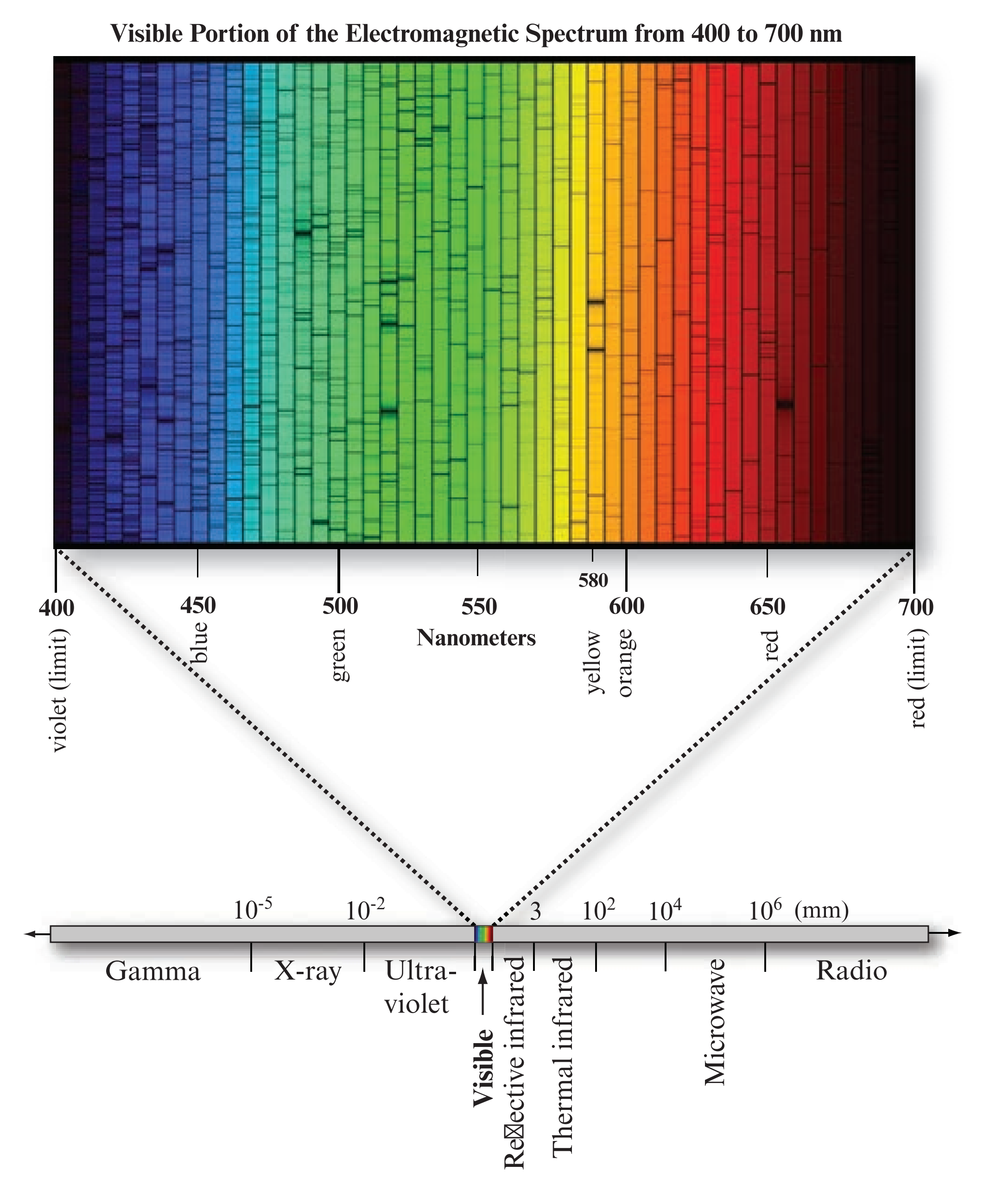


FIGURE 6–5 The portion of the electromagnetic spectrum visible by human beings is generally considered to be from 400 to 700 nm consisting of blue (400 – 500 nm), green (500 – 600 nm), and red (600 – 700 nm) light (Ready, 2014). The image of the solar spectrum was observed with the Fourier Transform Spectrometer at the McMath-Pierce Solar Facility at Kitt Peak National Observatory, near Tucson, AZ. Each of the 50 slices covers 60 angstroms and wavelength increases from top to bottom along each slice. Note that it contains relatively narrow, dark Fraunhofer lines named for the German physicist Joseph von Fraunhofer (1787–1826) who "mapped" the location of over 570 of them in the solar spectrum. Later scientists discovered that the dark lines are caused by chemical element absorption in the upper layers of the Sun and some absorption by oxygen molecules in the Earth's atmosphere (N. A. Sharp, NOAO/NSO/Kitt Peak FTS/AURA/NSF).

As mentioned in Chapter 1, in remote sensing research we often specify a particular region of the electromagnetic spectrum (e.g., red light) by identifying a beginning and ending wavelength (or frequency) and then attaching a description. This wavelength (or frequency) interval in the electromagnetic spectrum is commonly referred to as a *band*, *channel*, or *region*. The major subdivisions of visible light are presented diagrammatically in Figures 6-5, 6-6, and 6-7 and summarized in Table 6-2. For example, we generally think of visible light as being composed of energy in the blue (0.4 – 0.5 μm), green (0.5 – 0.6 μm), and red (0.6 – 0.7 μm) bands of the electromagnetic spectrum. Similarly, reflected near-infrared energy in the region from 0.7 to 1.3 μm is

TABLE 6–2 Methods of describing the color spectrum (after Nassau, 1983; 1984).

Color	Wavelength Descriptions				Energy Descriptions	
	Angstrom (A)	Nanometer (nm)	Micrometer (μm)	Frequency Hz (× 1,014)	Wave Number (ψ cm^{-1})	Electron Volt (eV)
Ultraviolet, sw	2,537	254	0.254	11.82	39,400	4.89
Ultraviolet, lw	3,660	366	0.366	8.19	27,300	3.39
Violet (limit)	4,000	400	0.40	7.50	25,000	3.10
Blue	4,500	450	0.45	6.66	22,200	2.75
Green	5,000	500	0.50	6.00	20,000	2.48
Green	5,500	550	0.55	5.45	18,200	2.25
Yellow	5,800	580	0.58	5.17	17,240	2.14
Orange	6,000	600	0.60	5.00	16,700	2.06
Red	6,500	650	0.65	4.62	15,400	1.91
Red (limit)	7,000	700	0.70	4.29	14,300	1.77
Infrared, near	10,000	1,000	1.00	3.00	10,000	1.24
Infrared, far	300,000	30,000	30.00	0.10	333	0.04

commonly used to expose black-and-white and color-infrared-sensitive film.

The middle-infrared region (often referred to as the short wavelength infrared, SWIR) includes energy with a wavelength of 1.3 to 3 μm. The thermal infrared region has two very useful bands at 3 to 5 μm and 8 to 14 μm. The microwave portion of the spectrum consists of much longer wavelengths (1 mm–1 m). The radio-wave portion of the spectrum may be subdivided into UHF, VHF, radio (HF), LF, and ULF frequencies.

The spectral resolution of most remote sensing systems is described in terms of bands of the electromagnetic spectrum. The band specifications for several of the most important remote sensing systems were summarized in Chapter 2.

Electromagnetic energy may be described not only in terms of wavelength and frequency but also in photon energy units such as joules (J) and electron volts (eV), as shown in Figure 6-8. Important mass, energy, and power conversions are summarized in Table 6-3.

The Particle Model: Radiation from Atomic Structures

In *Opticks* (1704), Sir Isaac Newton stated that light was a stream of particles, or corpuscles, traveling in straight lines. He also knew that light had wavelike characteristics based on his work with glass plates. Nevertheless, during the 200 years before 1905, light was thought of primarily as a smooth and continuous wave. Albert Einstein (1879–1955) found that when light interacts with electrons, it has a different character. He concluded that when light interacts with matter, it behaves as though it is composed of many individual bodies called *photons*, which carry particle-like properties such as energy and momentum. As a result, most physicists today would answer the question, “What is light?” by saying that light is a *particular* kind of matter. Thus, we sometimes describe electromagnetic energy in terms of its wavelike properties. But when the energy interacts with matter, it is useful to describe it as discrete packets of energy, or *quanta*. It is practical to review how electromagnetic energy is generated at the atomic level; this provides insight as to how light interacts with matter.

Blackbody Radiation Curves

6,000 K Sun
3,000 K Tungsten filament
800 K Red-hot object
300 K Earth
195 K Dry ice
79 K Liquid air
Visible Light
Relative Radiated Energy
Wavelength, μm

FIGURE 6–6 Blackbody radiation curves for several objects, including the Sun and the Earth, which approximate 6,000 K and 300 K blackbodies, respectively. The area under each curve may be summed to compute the total radiant energy (M_λ) exiting each object (Equation 6.4). Thus, the Sun produces more radiant exitance than the Earth because its temperature is greater. As the temperature of an object increases, its dominant wavelength (λ_{max}) shifts toward the shorter wavelengths of the spectrum.

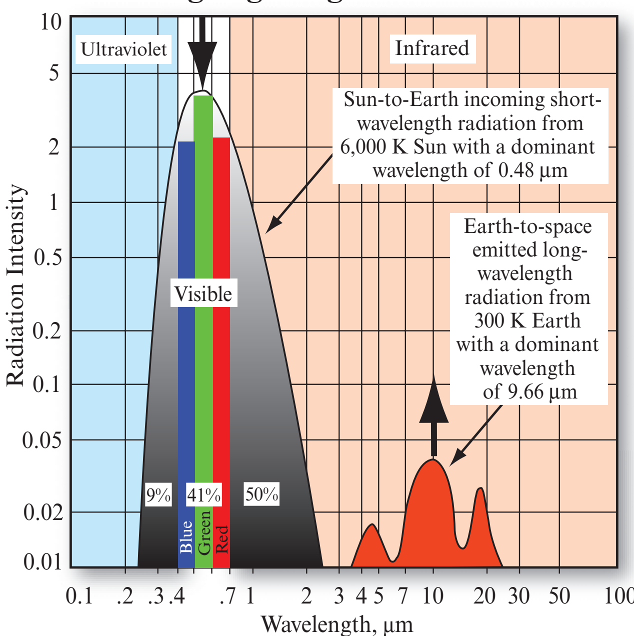


FIGURE 6–7 The Sun approximates a 6,000 K blackbody with a dominant wavelength of about 0.48 μm. The Earth approximates a 300 K blackbody with a dominant wavelength of about 9.66 μm. The 6,000 K Sun produces approximately 41% of its energy in the visible region from 0.4 to 0.7 μm (blue, green, and red light). The other 59% of the energy is in wavelengths shorter than blue light (<0.4 μm) and longer than red light (>0.7 μm). Our eyes are only sensitive to light from 0.4 to 0.7 μm (after Strahler and Strahler, 1989). Fortunately, it is possible to make remote sensor detectors sensitive to energy in these nonvisible regions of the spectrum.

Electrons are the tiny negatively charged particles that move around the positively charged nucleus of an atom (Figure 6-9). Atoms of different substances are made up of varying numbers of electrons arranged in different ways. The interaction between the positively charged nucleus and the negatively charged electron keeps the electron in orbit. While its orbit is not explicitly fixed, each electron's motion is restricted to a definite range from the nucleus. The allowable orbital paths of electrons moving around an atom might be thought of as energy classes or levels (Figure 6-9a). In order for an electron to climb to a higher class, work must be performed. However, unless an amount of energy is available to move the electron up at least one energy level, it will accept no work. If a sufficient amount of energy is received, the electron will jump to a new level and the atom is said to be *excited* (Figure 6-9b). Once an electron is in a higher orbit, it possesses potential energy. After about 10^{-8} seconds, the electron falls back to the atom's lowest empty energy level or orbit and gives off radiation (Figure 6-9c). The wavelength of radiation given off is a function of the amount of work done on the atom, i.e., the quantum of energy it absorbed to cause the electron to become excited and move to a higher orbit.

Electron orbits are like the rungs of a ladder. Adding energy moves the electron up the energy ladder; emitting energy moves it down. However, the energy ladder differs from an ordinary ladder in that its rungs are unevenly spaced. This means that the energy an electron needs to absorb, or to give up, in order to jump from one orbit to the next may not be the same as the energy change needed for some other step. Furthermore, an

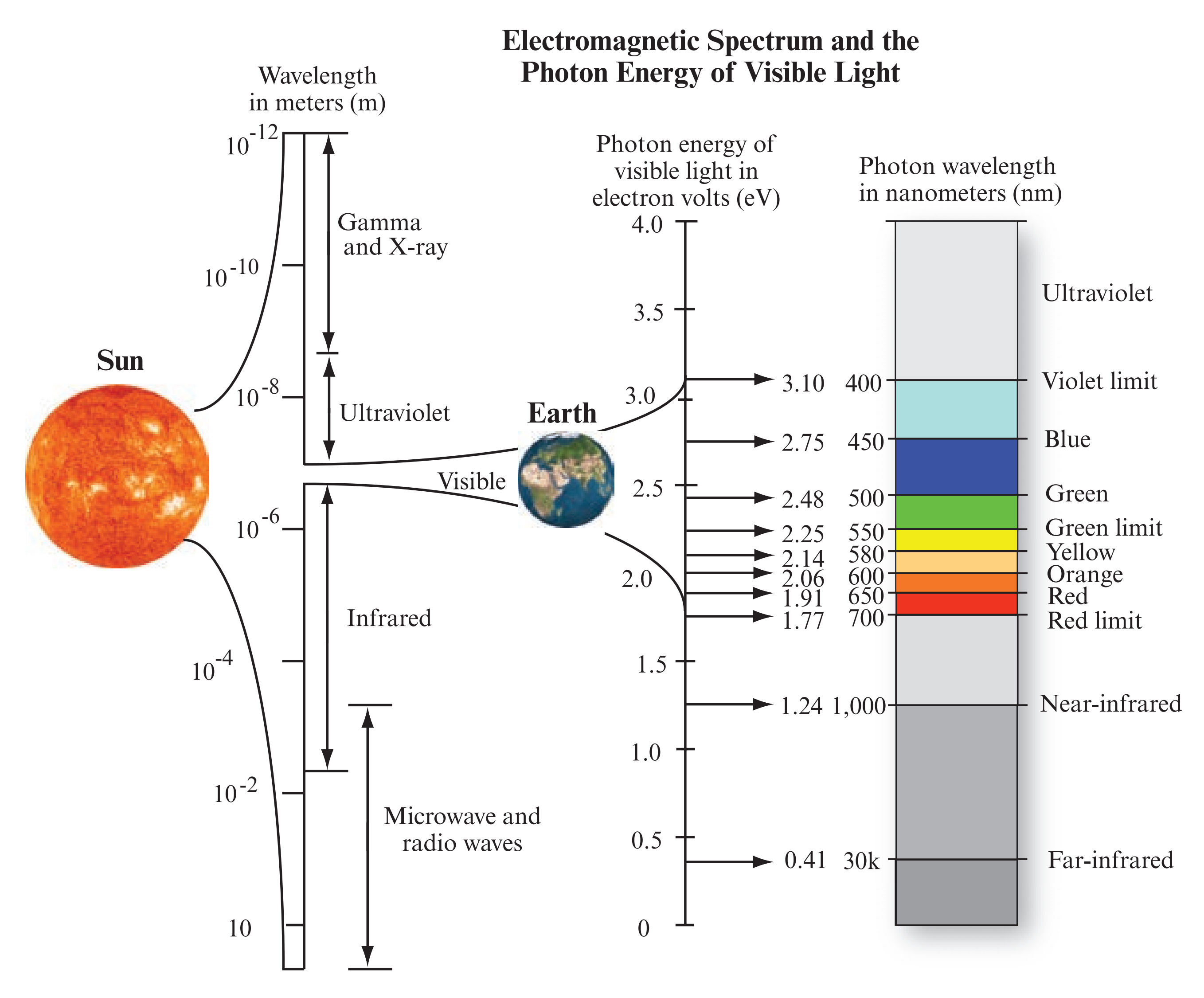


FIGURE 6–8 The electromagnetic spectrum and the photon energy of visible light. The Sun produces a continuous spectrum of energy from gamma rays to radio waves that continually bathe the Earth in energy. The visible portion of the spectrum may be measured using wavelength (measured in micrometers or nanometers, i.e., μm or nm) or electron volts (eV). All of the units are interchangeable.

electron does not necessarily use consecutive rungs. Instead, it follows what physicists call *selection rules.* In many cases, an electron uses one sequence of rungs as it climbs the ladder and another sequence as it descends (Nassau, 1983). The energy that is left over when the electrically charged electron moves from an excited state (Figure 6-9b) to a de-excited state (Figure 6-9c) is emitted by the atom as a packet of electromagnetic radiation, a particle-like unit of light called a *photon*. Every time an electron jumps from a higher to a lower energy level, a photon moves away at the speed of light.

Somehow an electron must disappear from its original orbit and reappear in its destination orbit without ever having to traverse any of the positions in between. This process is called a *quantum leap* or *quantum jump.* If the electron leaps from its highest excited state to the ground state in a single leap, it will emit a single photon of energy. It is also possible for the electron to leap from an excited orbit to the ground state in a series of jumps (e.g., from 4 to 2 to 1). If it takes two leaps to get to the ground state, then each of these jumps will emit photons of somewhat less energy. The energies emitted in the two jumps must sum to the total of the single large jump.

Niels Bohr (1885–1962) and Max Planck recognized the discrete nature of exchanges of radiant energy and proposed the *quantum theory* of electromagnetic radiation. This theory states that energy is transferred in discrete packets called quanta or photons, as discussed. The relationship between the frequency of radiation expressed by wave theory and the quantum is:

$$Q = h\nu \tag{6.6}$$

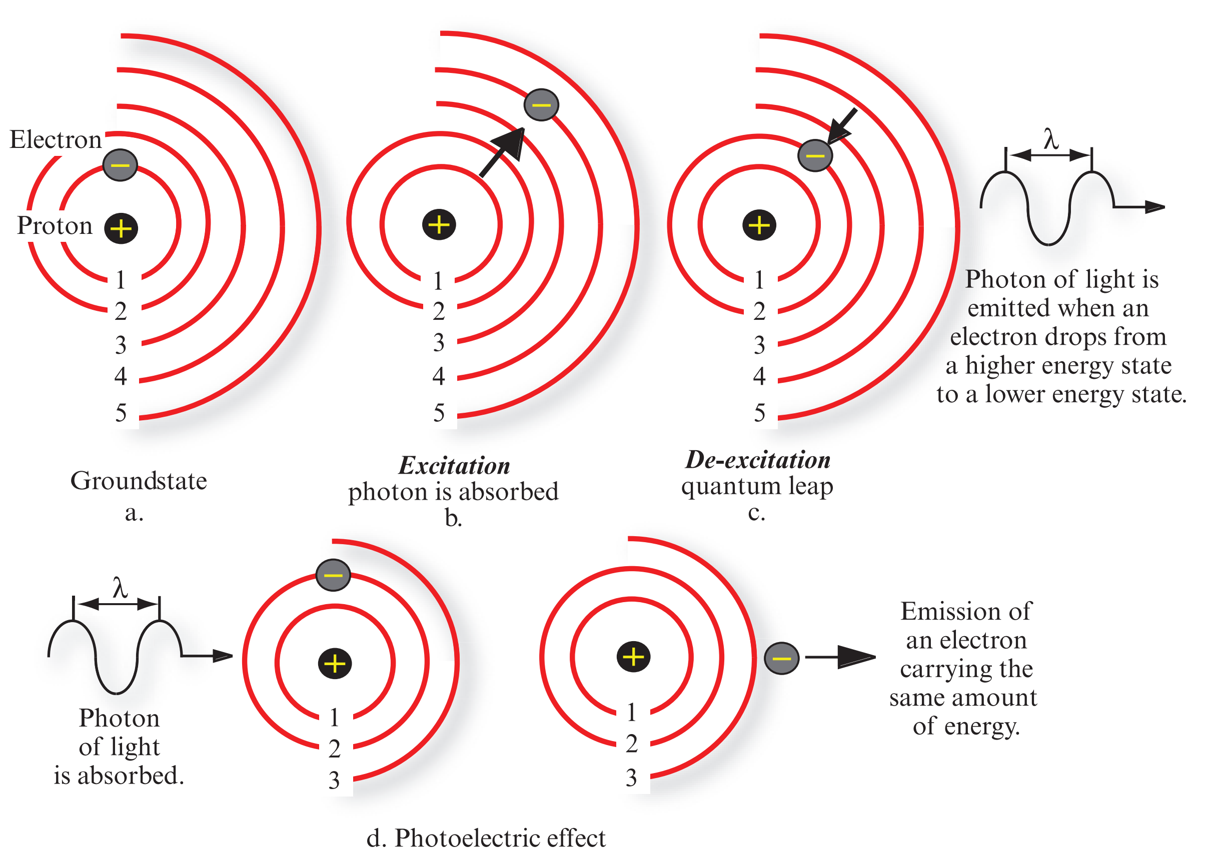


FIGURE 6–9 a–c) A photon of electromagnetic energy is emitted when an electron in an atom or molecule drops from a higher-energy state to a lower-energy state. The light emitted (i.e., its wavelength) is a function of the changes in the energy levels of the outer, valence electron. For example, yellow light is produced from a sodium vapor lamp in Figure 6-10. d) Matter can also be subjected to such high temperatures that electrons, which normally move in captured, nonradiating orbits, are broken free. When this happens, the atom remains with a positive charge equal to the negatively charged electron that escaped. The electron becomes a free electron, and the atom is called an ion. If another free electron fills the vacant energy level created by the free electron, then radiation from all wavelengths is produced, i.e., a continuous spectrum of energy. The intense heat at the surface of the Sun produces a continuous spectrum in this manner.

where Q is the energy of a quantum measured in joules, h is the Planck constant (6.626×10^{-34} J s), and ν is the frequency of the radiation. Referring to Equation 6.3, we can multiply the equation by h/h, or 1, without changing its value:

$$\lambda = \frac{hc}{h\nu}. \qquad (6.7)$$

By substituting Q for $h\nu$ (from Equation 6.6), we can express the wavelength associated with a quantum of energy as:

$$\lambda = \frac{hc}{Q} \qquad (6.8)$$

or

$$Q = \frac{hc}{\lambda}. \qquad (6.9)$$

Thus, we see that the energy of a quantum is inversely proportional to its wavelength, i.e., the longer the wavelength, the lower its energy content. This inverse relationship is important to remote sensing because it suggests that it is more difficult to detect the longer-wavelength energy being emitted at thermal infrared wavelengths than those at shorter visible wavelengths. In fact, it might be necessary to have the sensor look at or dwell longer on the parcel of ground if we are trying to measure the longer wavelength energy.

Substances have color because of differences in their energy levels and the selection rules. For example, consider energized sodium vapor that produces a bright yellow light that is used in some street lamps. When a sodium vapor lamp is turned on, several thousand volts of electricity energize the vapor. The outermost

TABLE 6–3 Mass, energy, and power conversions.

Conversion from English to SI Units		
To get:	**Multiply:**	**By:**
newtons[a]	pounds	4.448
joules[b]	BTUs[c]	1055
joules	calories[d]	4.184
joules	kilowatt-hours[e]	3.6×10^6
joules	foot-pounds[f]	1.356
joules	horsepower[g]	745.7
Conversion from SI to English Units		
To get:	**Multiply:**	**By:**
BTUs	joules	0.00095
calories	joules	0.2390
kilowatt-hours	joules	2.78×10^{-7}
foot-pounds	joules	0.7375
horsepower	watts	0.00134

[a]newton: force needed to accelerate a mass of 1 kg by 1 m s^{-2}
[b]joule: a force of 1 newton acting through 1 meter.
[c]British thermal unit, or BTU: energy required to raise the temperature of 1 pound of water by 1 degree Fahrenheit.
[d]calorie: energy required to raise the temperature of 1 kilogram of water by 1 degree Celsius.
[e]kilowatt-hour: 1,000 joules per second for 1 hour.
[f]foot-pound: a force of 1 pound acting through 1 foot.
[g]horsepower: 550 foot-pounds per second.

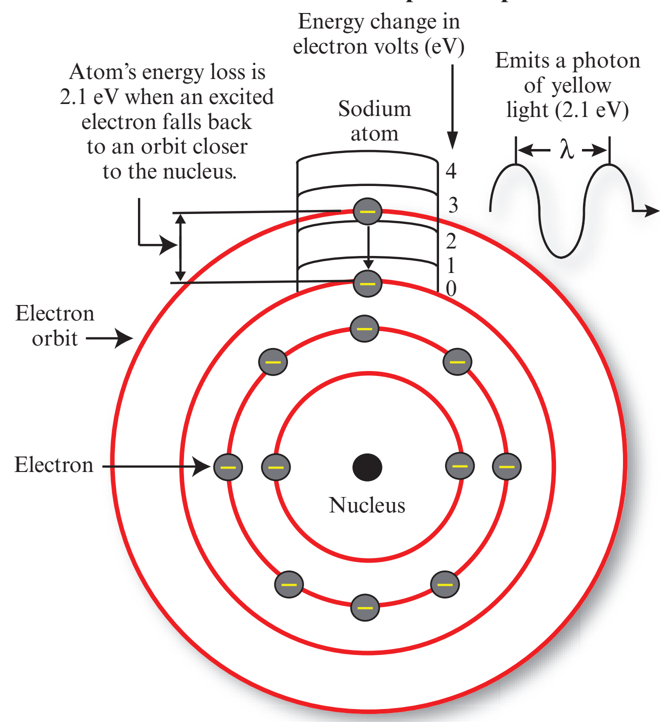


FIGURE 6–10 Creation of light from atomic particles in a sodium vapor lamp. After being energized by several thousand volts of electricity, the outermost electron in each energized atom of sodium vapor climbs to a high rung on the energy ladder and then returns down the ladder in a pre-d t able fashio The las two rungs in the descent are 2.1 eV apart. This produces a photon of yellow light, which has 2.1 eV of energy (see Table 6-2).

electron in each energized atom of sodium vapor climbs to a higher rung on the energy ladder and then returns down the ladder in a certain sequence of rungs, the last two of which are 2.1 eV apart (Figure 6-10). The energy released in this last leap appears as a photon of yellow light with a wavelength of 0.58 μm with 2.1 eV of energy (Nassau, 1983).

Matter can be heated to such high temperatures that electrons that normally move in captured, nonradiating orbits break free. This is called the *photoelectric effect* (Figure 6-9d). When this happens, the atom remains with a positive charge equal to the negatively charged electron that escaped. The electron becomes a free electron and the atom is called an *ion*. In the ultraviolet and visible (blue, green, and red) parts of the electromagnetic spectrum, radiation is produced by changes in the energy levels of the outer valence electrons. The wavelengths of energy produced are a function of the particular orbital levels of the electrons involved in the excitation process. If the atoms absorb enough energy to become ionized and if a free electron drops in to fill the vacant energy level, then the radiation given off is unquantized and a *continuous spectrum* is produced rather than a band or a series of bands. Every encounter of one of the free electrons with a positively charged nucleus causes rapidly changing electric and magnetic fields, so that radiation at all wavelengths is produced. The hot surface of the Sun is largely a *plasma* in which radiation of all wavelengths is produced. As previously shown in Figure 6-8, the spectrum of a plasma like the Sun is continuous.

In atoms and molecules, the changes in electron orbits produce the shortest wavelength radiation, molecule

vibrational motion changes produce near-infrared and/or middle-infrared energy, and rotational motion changes produce long-wavelength infrared or microwave radiation.



Atmospheric Energy–Matter Interactions

Radiant energy is the capacity of radiation within a spectral band to do work. Once electromagnetic radiation is generated, it is propagated through the Earth's atmosphere almost at the speed of light in a vacuum. Unlike a vacuum in which nothing happens, however, the atmosphere may affect not only the speed of radiation but also its wavelength, its intensity, and its spectral distribution. The electromagnetic radiation may also be diverted from its original direction due to refraction.

Refraction

The speed of light in a vacuum is 3×10^8 m/s. When electromagnetic radiation encounters substances of different density, like air and water, refraction may take place. *Refraction* refers to the bending of light when it passes from one medium to another of different density. Refraction occurs because the media are of differing densities and the speed of EMR is different in each. The *index of refraction* (n) is a measure of the optical density of a substance. This index is the ratio of the speed of light in a vacuum, c, to the speed of light in a substance such as the atmosphere or water, c_n:

$$n = \frac{c}{c_n}. \tag{6.10}$$

The speed of light in a substance can never reach the speed of light in a vacuum. Therefore, its index of refraction must always be greater than 1. For example, the index of refraction for the atmosphere is 1.0002926 and 1.33 for water. Light travels more slowly through water because of water's higher density.

Refraction can be described by Snell's law, which states that for a given frequency of light (we must use frequency since, unlike wavelength, it does not change when the speed of light changes), the product of the index of refraction and the sine of the angle between the ray and a line normal to the interface is constant:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{6.11}$$

From Figure 6-11 we can see that a nonturbulent atmosphere can be thought of as a series of layers of gases, each with a slightly different density. Any time

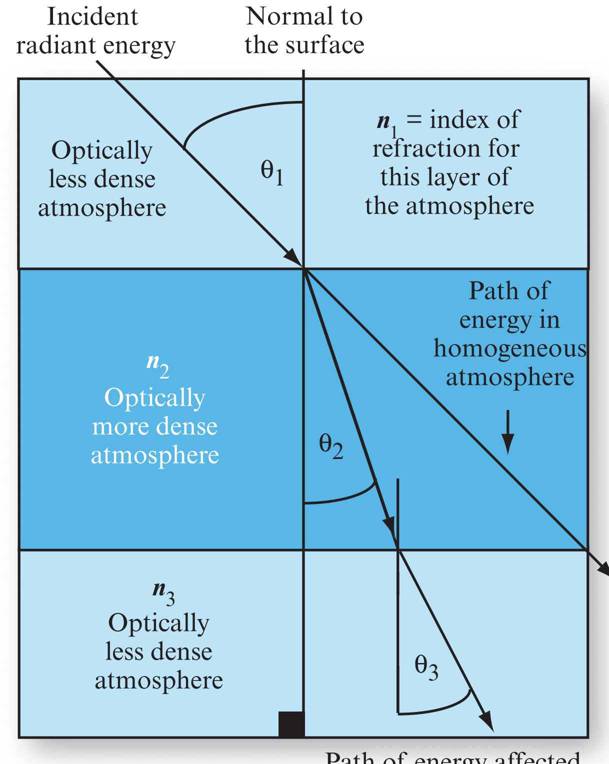


FIGURE 6–11 Refraction in three nonturbulent atmospheric layers. The incident energy is bent from its normal trajectory as it travels from one atmospheric layer to another. Snell's law can be used to predict how much bending will take place, based on a knowledge of the angle of incidence (θ) and the index of refraction of each atmospheric layer, n_1, n_2, n_3.

energy is propagated through the atmosphere for any appreciable distance at any angle other than vertical, refraction occurs.

The amount of refraction is a function of the angle made with the vertical (θ), the distance involved (in the atmosphere the greater the distance, the more changes in density), and the density of the air involved (air is usually more dense near sea level). Serious errors in location due to refraction can occur in images formed from energy detected at high altitudes or acute angles. However, these location errors are predictable by Snell's law and thus can be removed. Notice that

$$\sin\theta_2 = \frac{n_1 \sin\theta_1}{n_2}. \tag{6.12}$$

Therefore, if an analyst knows the index of refraction of medium n_1 and n_2 and the angle of incidence of the energy to medium n_1, it is possible to predict the

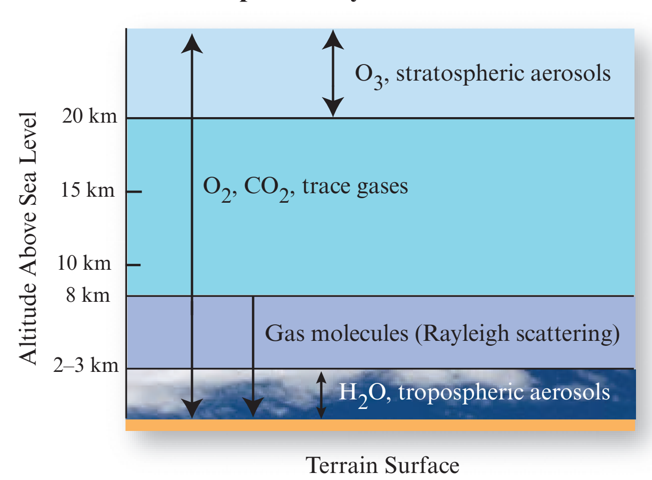
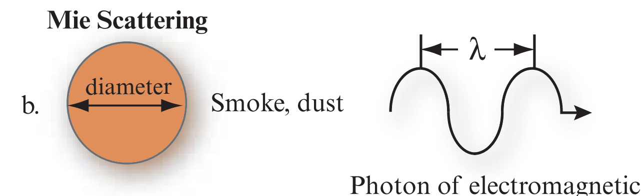


FIGURE 6–12 Major subdivisions of the atmosphere and the types of molecules and aerosols found in each layer (based on Miller and Vermote, 2002).



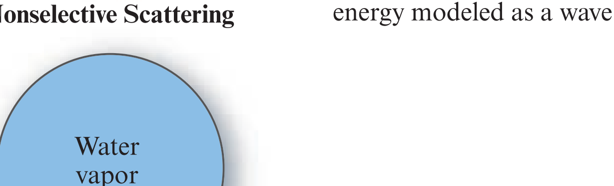


FIGURE 6–13 Type of scattering is a function of 1) the wavelength of the incident radiant energy, and 2) the size of the gas molecule, dust particle, and/or water vapor droplet encountered.

amount of refraction that will take place ($\sin \theta_2$) in medium n_2 using trigonometric relationships. Interestingly, most image analysts never concern themselves with computing the index of refraction.

Scattering

One very serious effect of the atmosphere is the scattering of radiation by atmospheric particles. *Scattering* differs from reflection in that the direction associated with scattering is unpredictable, whereas the direction of reflection (to be defined shortly) is predictable. There are essentially three types of scattering: Rayleigh, Mie, and nonselective scattering. Major subdivisions of the atmosphere and the types of molecules and aerosols found in each layer are shown in Figure 6-12 (Miller and Vermote, 2002). The relative size of the wavelength of the incident electromagnetic radiation, the diameter of the gases, water vapor, and/or dust with which the energy interacts, and the type of scattering that should occur are summarized in Figure 6-13.

Rayleigh scattering (often referred to as *molecular scattering*) occurs when the effective diameter of the matter (usually air molecules such as oxygen and nitrogen in the atmosphere) are many times smaller (usually < 0.1) than the wavelength of the incident electromagnetic radiation (Figure 6-13a). Rayleigh scattering is named after the English physicist Lord Rayleigh, who offered the first coherent explanation for it. All scattering is accomplished through absorption and re-emission of radiation by atoms or molecules in the manner previously described in the section on radiation from atomic structures. It is impossible to predict the direction in which a specific atom or molecule will emit a photon, hence scattering. The energy required to excite an atom is associated with powerful short-wavelength, high-frequency radiation.

The amount of Rayleigh scattering in the atmosphere in optical wavelengths (0.4–0.7 μm) may be computed using the Rayleigh scattering cross-section (τ_m) algorithm (Cracknell and Hayes, 1993):

$$\tau_m = \frac{8\pi^3(n^2-1)^2}{(3N^2\lambda^4)} \tag{6.13}$$

where n = refractive index, N = number of air molecules per unit volume, and λ = wavelength. The amount of scattering is inversely related to the fourth power of the radiation's wavelength. For example, ultraviolet light at 0.3 μm is scattered approximately 16 times more than red light at 0.6 μm, i.e., $(0.6/0.3)^4 = 16$. Blue light at 0.4 μm is scattered about 5 times more than red light at 0.6 μm, i.e., $(0.6/0.4)^4 = 5.06$. The amount of Rayleigh scattering expected throughout the visible part of the spectrum (0.4 – 0.7 μm) is shown in Figure 6-14.

Most Rayleigh scattering by gas molecules takes place in the atmosphere 2 to 8 km above the ground (Figure 6-12). Rayleigh scattering is responsible for the blue appearance of the sky. The shorter violet and blue

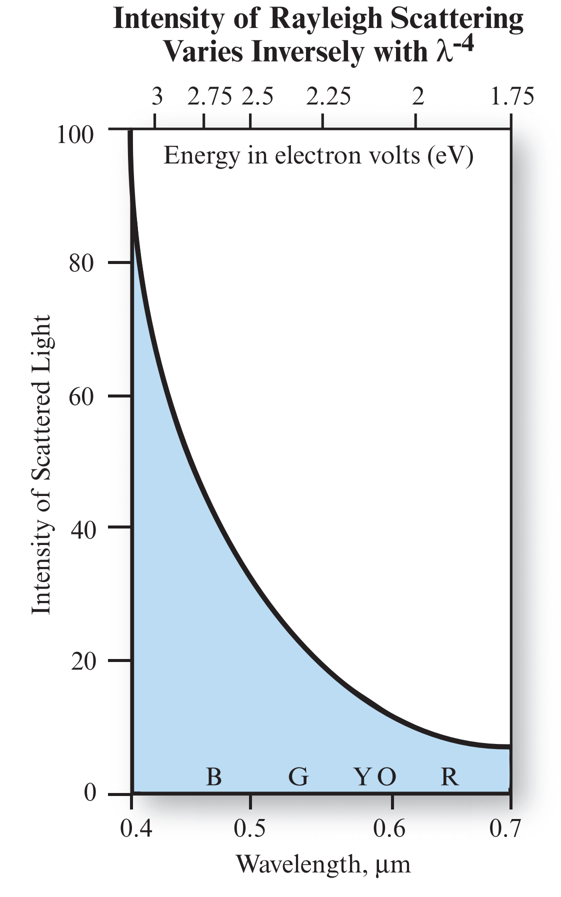


FIGURE 6–14 The intensity of Rayleigh scattering varies inversely with the fourth power of the wavelength (λ^{-4}).

wavelengths are more efficiently scattered than the longer orange and red wavelengths. When we look up on a cloudless day and admire the blue sky, we are witnessing the preferential scattering of the short-wavelength sunlight. Rayleigh scattering is also responsible for red sunsets. Since the atmosphere is a thin shell of gravitationally bound gas surrounding the solid Earth, sunlight must pass through a longer slant path of air at sunset (or sunrise) than at noon. Since the violet and blue wavelengths are scattered even more during their now longer path through the air than when the Sun is overhead, what we see when we look toward the sunset is the residue—the wavelengths of sunlight that are hardly scattered away at all, especially the oranges and reds.

Mie scattering (sometimes referred to as nonmolecular or aerosol particle scattering) takes place in the lower 4.5 km of the atmosphere, where there may be many essentially spherical particles present with diameters approximately equal to the size of the wavelength of the incident energy (Figure 6-13b). The actual size of the particles may range from 0.1 to 10 times the wavelength of the incident energy. For visible light, the main scattering agents are dust and other particles ranging from a few tenths of a micrometer to several micrometers in diameter. The amount of scatter is greater than Rayleigh scatter, and the wavelengths scattered are longer. The greater the amount of smoke and dust particles in the atmospheric column, the more that violet and blue light will be scattered away and only the longer wavelength yellow, orange, and red light will reach our eyes. Pollution also contributes to beautiful sunsets and sunrises.

Nons elective scattering takes place in the lowest portions of the atmosphere where there are particles >10 times the wavelength of the incident electromagnetic radiation (Figure 6-13c). This type of scattering is nonselective, i.e., all wavelengths of light are scattered, not just blue, green, or red. Thus, the water droplets and ice crystals that make up clouds and fog banks scatter all wavelengths of visible light equally well, causing the cloud to appear white. Nonselective scattering of approximately equal proportions of blue, green, and red light always appears as white light to the casual observer. This is the reason why putting our automobile high beams on in fog only makes the problem worse as we nonselectively scatter even more light into our visual field of view.

Scattering is a very important consideration in remote sensing investigations. It can severely reduce the information content of remotely sensed data to the point that the imagery loses contrast and it becomes difficult to differentiate one object from another.

Absorption

Absorption is the process by which radiant energy is absorbed and converted into other forms of energy. The absorption of the incident radiant energy may take place in the atmosphere or on the terrain. An *absorption band* is a range of wavelengths (or frequencies) in the electromagnetic spectrum within which radiant energy is absorbed by a substance. The effects of water (H_2O), carbon dioxide (CO_2), oxygen (O_2), ozone (O_3), and nitrous oxide (N_2O) on the transmission of light through the atmosphere are summarized in Figure 6-15a. The cumulative effect of the absorption by the various constituents can cause the atmosphere to “close down” completely in certain regions of the spectrum. This is bad for remote sensing because no energy is available to be sensed. Conversely, in the visible portion of the spectrum (0.4 – 0.7 μm), the atmosphere does not absorb all of the incident energy but transmits it rather effectively. Portions of the spectrum that transmit radiant energy effectively are called *atmospheric windows*.

Absorption occurs when incident energy of the same frequency as the resonant frequency of an atom or

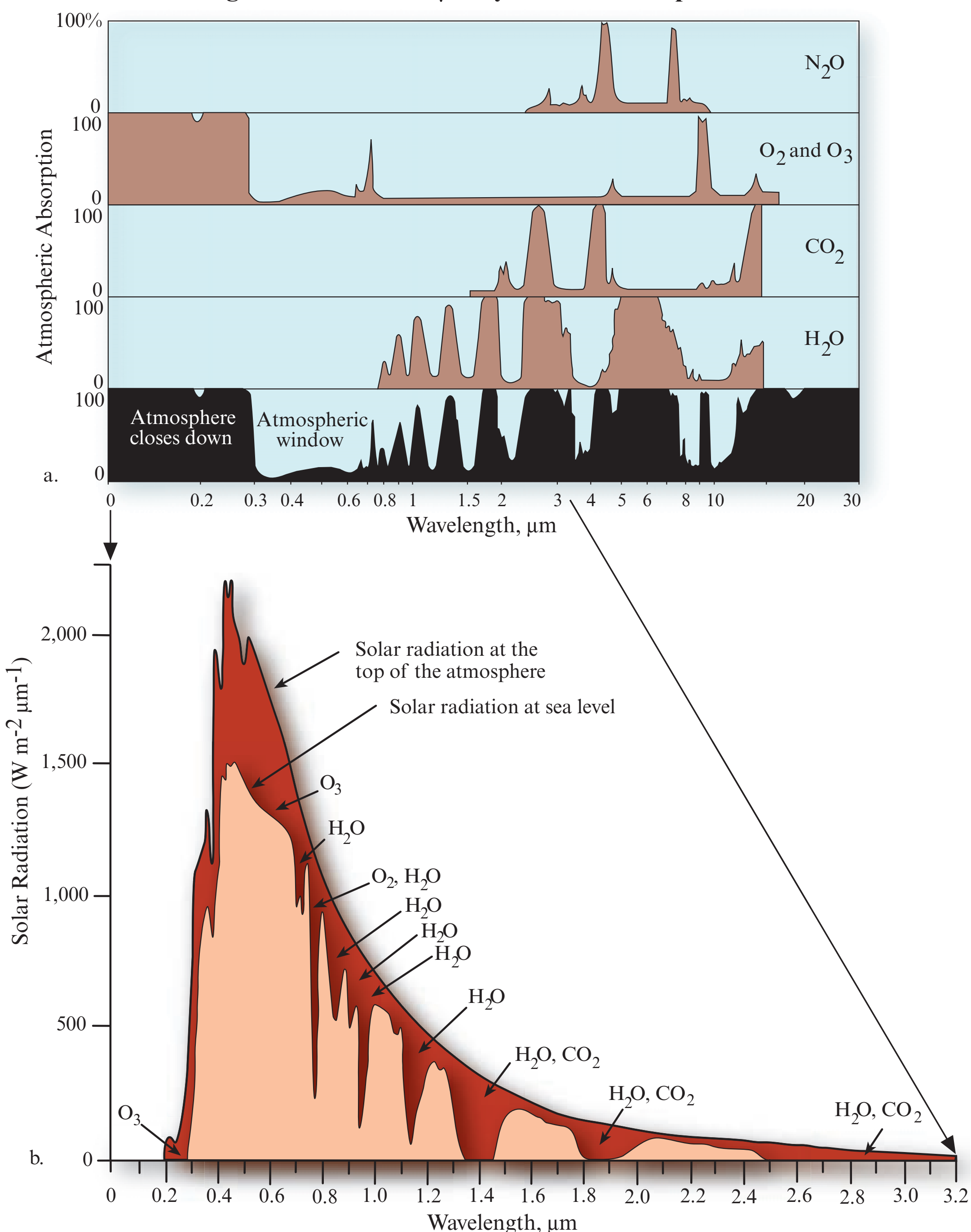


FIGURE 6–15 a) The absorption of the Sun's incident electromagnetic energy in the region from 0.1 to 30 μm by various atmospheric gases. The first four graphs depict the absorption characteristics of N_2O, O_2 and O_3, CO_2, and H_2O, while the final graphic depicts the cumulative result of all these constituents being in the atmosphere at one time. The atmosphere essentially "closes down" in certain portions of the spectrum while "atmospheric windows" exist in other regions that transmit incident energy effectively to the ground. It is within these windows that remote sensing systems must function. b) The combined effects of atmospheric absorption, scattering, and reflectance reduce the amount of solar irradiance reaching the Earth's surface at sea level (based on Slater, 1980).

molecule is absorbed, producing an excited state. If instead of reradiating a photon of the same wavelength, the energy is transformed into heat motion and is subsequently reradiated at a longer wavelength, absorption occurs. When dealing with a medium like air, absorption and scattering are frequently combined into an *extinction coefficient* (Konecny, 2014). Transmission is inversely related to the extinction coefficient

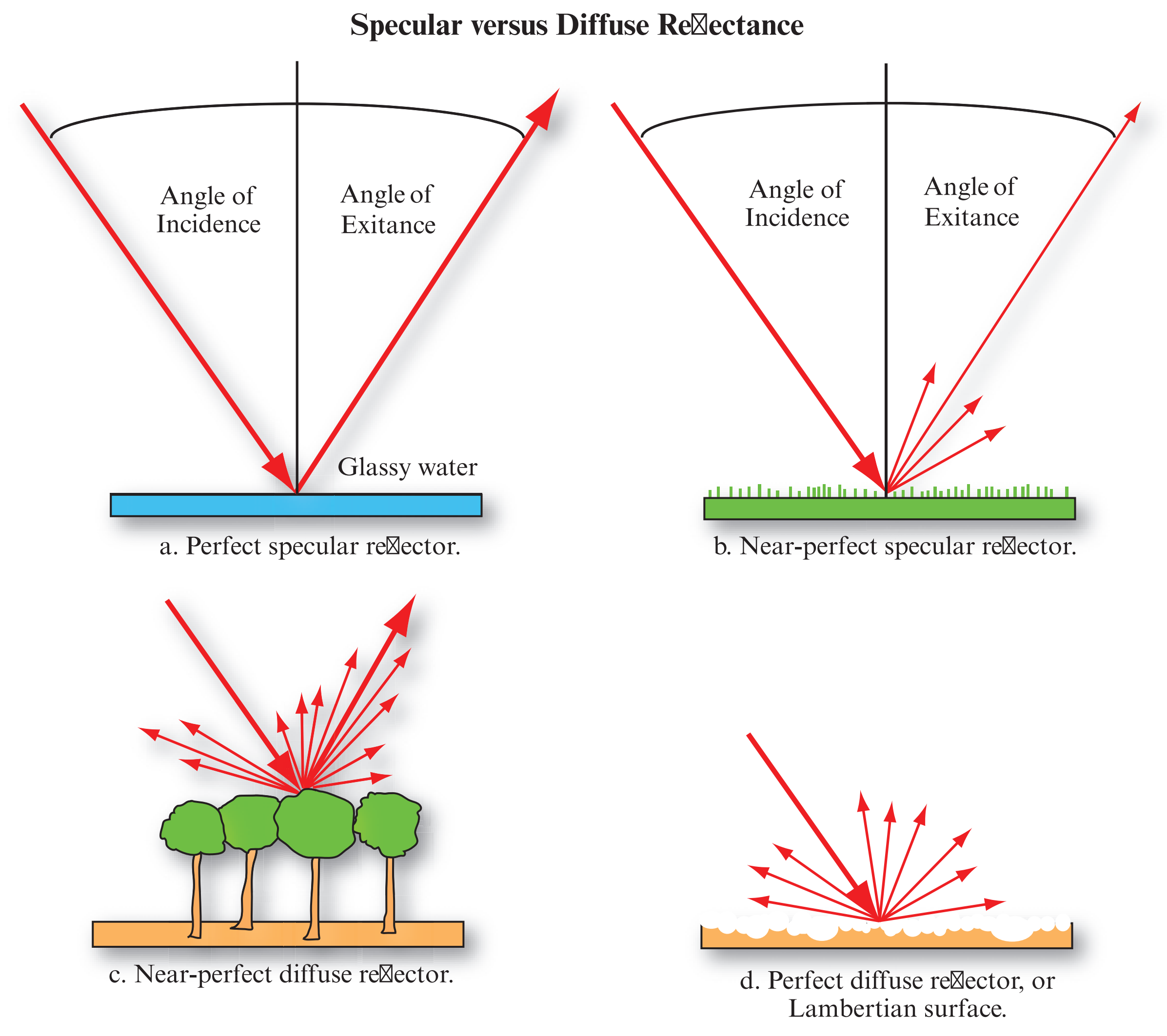


FIGURE 6–16 The nature of specular and diffuse reflectance.

times the thickness of the layer. Certain wavelengths of radiation are affected far more by absorption than by scattering. This is particularly true of infrared and wavelengths shorter than visible light. The combined effects of atmospheric absorption, scattering, and reflectance (from cloud tops) can dramatically reduce the amount of solar radiation reaching the Earth's surface at sea level, as shown in Figure 6-15b.

Energy reaching the remote sensing system must pass through the atmosphere twice. Therefore, it is common to identify two *atmospheric transmission coefficients*: one for energy coming into the atmosphere (T_{θ_o}) at an incidence angle related to the source of the energy (e.g., the Sun) and one for the atmosphere that the Earth surface–reflected or emitted energy must pass through to reach the remote sensor system (T_{θ_v}). We will see how important these two parameters are in the discussion on atmospheric correction later in this chapter.

Reflectance

Reflectance is the process whereby radiation "bounces off" an object like the top of a cloud, a water body, or the terrestrial Earth. Actually, the process is more complicated, involving reradiation of photons in unison by atoms or molecules in a layer approximately one-half wavelength deep. Reflection exhibits fundamental characteristics that are important in remote sensing. First, the incident radiation, the reflected radiation, and a vertical to the surface from which the angles of incidence and reflection are measured all lie in the same plane. Second, the angle of incidence and the angle of reflection (exitance) are approximately equal, as shown in Figure 6-16.

There are various types of reflecting surfaces. *Specular reflection* occurs when the surface from which the radiation is reflected is essentially smooth (i.e., the average surface-profile height is several times smaller than the wavelength of radiation striking the surface). Several

TABLE 6–4 Radiometric concepts (based on Colwell, 1983; Vandergriff, 2014).

Name	Symbol	Units	Concept
Radiant energy	Q_λ	joules, J	Capacity of radiation within a specified spectral band to do work.
Radiant flux	Φ_λ	watts, W	Time rate of flow of energy onto, off of, or through a surface.
Radiant flux density at the surface **Irradiance** **Radiant exitance**	 E_λ M_λ	 watts per square meter, W m^{-2} watts per square meter, W m^{-2}	 Radiant flux incident upon a surface per unit area of that surface. Radiant flux leaving a surface per unit area of that surface.
Radiance (spectral)	L_λ	watts per square meter, per steradian, W m^{-2} sr^{-1}	Radiant intensity per unit of projected source area in a specified direction.
Hemispherical reflectance	ρ_λ	dimensionless	$\frac{\Phi_{\text{reflected}_\lambda}}{\Phi_{i_\lambda}}$
Hemispherical transmittance	τ_λ	dimensionless	$\frac{\Phi_{\text{transmitted}_\lambda}}{\Phi_{i_\lambda}}$
Hemispherical absorptance	α_λ	dimensionless	$\frac{\Phi_{\text{absorbed}_\lambda}}{\Phi_{i_\lambda}}$

features, such as very calm water bodies, act like *near-perfect specular reflectors* (Figure 6-16ab). If there are very few ripples on the surface, the incident energy will leave the water body at an angle equal and opposite to the incident energy (Pedrotti, 2014). We know this occurs from our personal experience. If we shine a flashlight at night on a tranquil pool of water, the light will bounce off the surface and into the trees across the way at an angle equal to and opposite from the incident radiation angle.

If the surface has a large surface height relative to the size of the wavelength of the incident energy, the reflected rays go in many directions, depending on the orientation of the smaller reflecting surfaces. This *diff useref lœtion* does not yield a mirror image, but instead produces diffused radiation (Figure 6-16c). White paper, white powders, and other materials reflect visible light in this diffuse manner. If the surface is so rough that there are no individual reflecting surfaces, then unpredictable scattering may occur. Lambert defined a perfectly diffuse surface; hence, the commonly designated *Lambertian surface* is one for which the radiant flux leaving the surface is constant for any angle of reflectance (Figure 6-16d).

A considerable amount of incident radiant flux from the Sun is reflected from the tops of clouds and other materials in the atmosphere. A substantial amount of this energy is reradiated back to space. As we shall see, the specular and diffuse reflection principles that apply to clouds also apply to the terrain.



Terrain Energy–Matter Interactions

The time rate of flow of energy onto, off of, or through a surface is called *radiant flux* (Φ) and is measured in watts (W) (Table 6-4). The characteristics of the radiant flux and what happens to it as it interacts with the Earth's surface is of critical importance in remote sensing. In fact, this is the fundamental focus of much remote sensing research. By carefully monitoring the exact nature of the incoming (incident) radiant flux in selective wavelengths and how it interacts with the terrain, it is possible to learn important information about the terrain.

Radiometric quantities have been identified that allow analysts to keep a careful record of the incident and exiting radiant flux (Table 6-4). We begin with the simple *radiation budget equation,* which states that the total amount of radiant flux in specific wavelengths (λ) incident to the terrain (Φ_{i_λ}) must be accounted for by evaluating the amount of radiant flux reflected from

the surface ($\Phi_{\text{reflected}_\lambda}$), the amount of radiant flux absorbed by the surface ($\Phi_{\text{absorbed}_\lambda}$), and the amount of radiant flux transmitted through the surface ($\Phi_{\text{transmitted}_\lambda}$):

$$\Phi_{i_\lambda} = \Phi_{\text{reflected}_\lambda} + \Phi_{\text{absorbed}_\lambda} + \Phi_{\text{transmitted}_\lambda} . \qquad (6.14)$$

It is important to note that these radiometric quantities are based on the amount of radiant energy incident to a surface from any angle in a hemisphere (i.e., a half of a sphere).

Hemispherical Reflectance, Absorptance, and Transmittance

Hemispherical reflectance (ρ_λ) is defined as the dimensionless ratio of the radiant flux reflected from a surface to the radiant flux incident to it (Table 6-4):

$$\rho_\lambda = \frac{\Phi_{\text{reflected}_\lambda}}{\Phi_{i_\lambda}} . \qquad (6.15)$$

Hemispherical transmittance (τ_λ) is defined as the dimensionless ratio of the radiant flux transmitted through a surface to the radiant flux incident to it:

$$\tau_\lambda = \frac{\Phi_{\text{transmitted}_\lambda}}{\Phi_{i_\lambda}} . \qquad (6.16)$$

Hemispherical absorptance (α_λ) is defined by the dimensionless relationship:

$$\alpha_\lambda = \frac{\Phi_{\text{absorbed}_\lambda}}{\Phi_{i_\lambda}} \qquad (6.17)$$

or

$$\alpha_\lambda = 1 - (r_\lambda + \tau_\lambda) . \qquad (6.18)$$

These definitions imply that radiant energy must be conserved, i.e., it is either returned back by reflection, transmitted through a material, or absorbed and transformed into some other form of energy inside the terrain. The net effect of absorption of radiation by most substances (e.g., plants, soil, rocks, water) is that the energy is converted into heat, causing a subsequent rise in the substance's temperature.

These radiometric quantities are useful for producing general statements about the spectral reflectance, absorptance, and transmittance characteristics of terrain features. In fact, if we take the simple hemispherical reflectance equation and multiply it by 100, we obtain an expression for percent reflectance ($\rho_{\lambda_\%}$):

$$\rho_{\lambda_\%} = \frac{\Phi_{\text{reflected}_\lambda}}{\Phi_{i_\lambda}} \times 100 , \qquad (6.19)$$

which is often used in remote sensing research to describe the spectral reflectance characteristics of various phenomena.

Examples of spectral percent reflectance curves for selected urban–suburban phenomena such as concrete, sand, sandy clay soil, red brick, Centipede grass, artificial turf, water, and asphalt shingles are shown in Figure 6-17. Spectral reflectance curves provide no information about the absorption and transmittance of the radiant energy. But because many of the sensor systems such as cameras and some multispectral scanners record only reflected energy, this information is still quite valuable and can form the basis for object identification and assessment. For example, it is clear from Figure 6-17 that the Centipede grass reflects only approximately 3 to 4% of the incident red radiant energy (0.6 – 0.7 μm) while reflecting up to approximately 35% of the incident near-infrared radiant flux (0.7 – 0.9 μm). If we wanted to discriminate between grass and artificial turf, the ideal portion of the spectrum to remotely sense in would be the near-infrared region because artificial turf reflects only about 5 percent of the incident near-infrared energy. This would cause a black-and-white infrared image of the terrain to display Centipede grass in bright tones and the artificial turf in darker tones. Please note how concrete reflects approximately 35% of the incident energy in the blue, green, and red parts of the spectrum, causing it to appear gray when humans view it a natural color photograph. This will be demonstrated in a subsequent section in this chapter.

Hemispherical reflectance, transmittance, and absorptance of radiometric quantities do not provide information about the exact amount of energy reaching a specific area on the ground from a specific direction or about the exact amount of radiant flux exiting the ground in a certain direction. Remote sensing systems can be located in space only at a single point in time, and they usually look only at a relatively small portion of the Earth at a single instant in time. Therefore, it is important to refine our radiometric measurement techniques so that more precise radiometric information can be extracted from the remotely sensed data. This requires the introduction of several radiometric quantities that provide progressively more precise radiometric information.

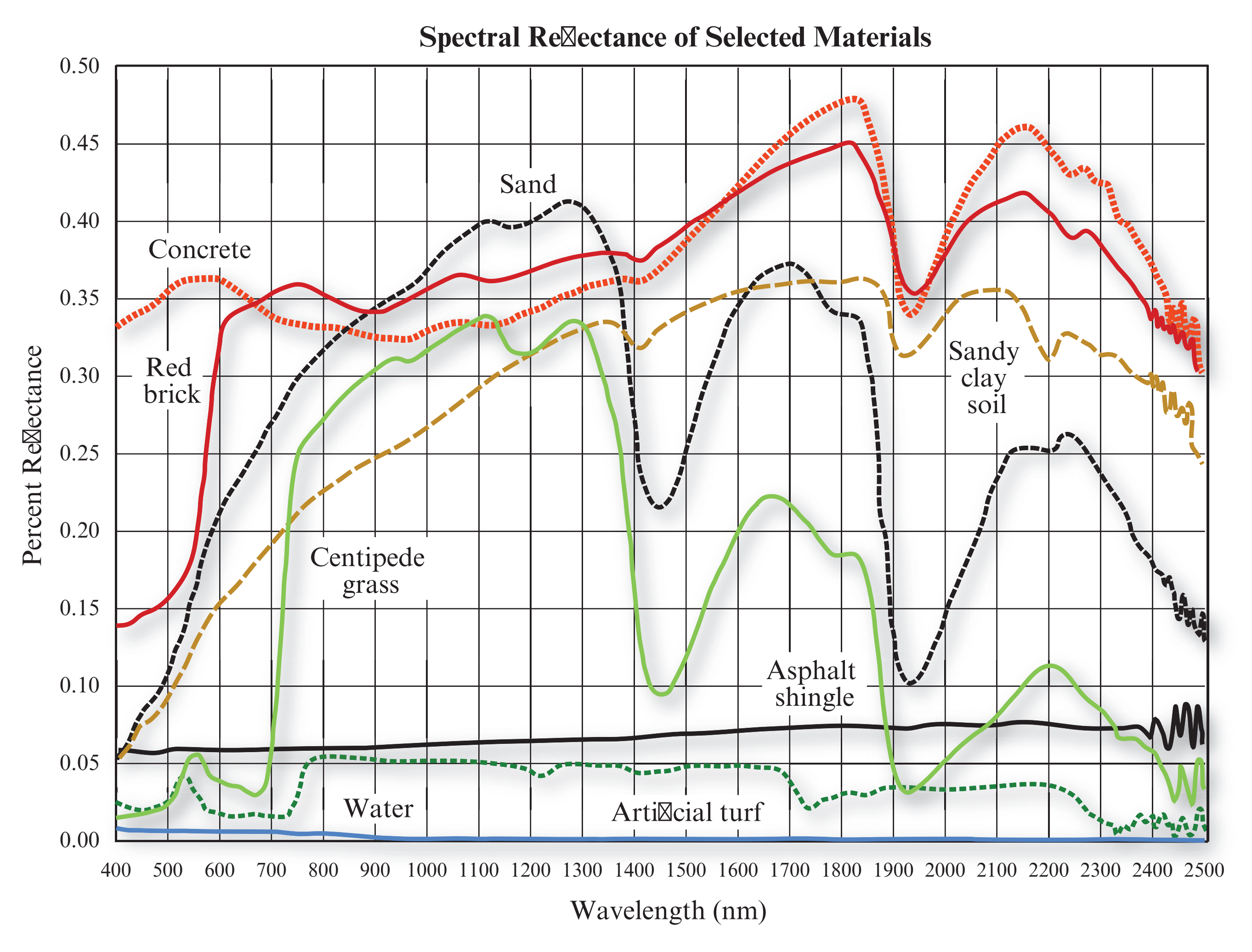


FIGURE 6–17 Spectral reflectance curves for selected urban–suburban phenomena in the region 400 – 2,500 nm.

Radiant Flux Density

A flat area (e.g., 1 × 1 m in dimension) being bathed in radiant flux (Φ) in specific wavelengths from the Sun is shown in Figure 6-18. The amount of radiant flux intercepted divided by the area of the plane surface is the average *radiant flux density.*

Irradiance and Exitance

The amount of radiant flux incident upon a surface per unit area of that surface is called *irradiance* (E_λ):

$$E_\lambda = \frac{\Phi_{i\lambda}}{A}. \qquad (6.20)$$

The amount of radiant flux leaving a surface per unit area of that surface is called *exitance* (M_λ):

$$M_\lambda = \frac{\Phi_{e\lambda}}{A}. \qquad (6.21)$$

Both quantities are usually measured in watts per meter squared (W m^{-2}). Although we do not have information on the direction of either the incoming or outgoing radiant energy (i.e., the energy can come and go at any angle throughout the entire hemisphere), we have now refined the measurement to include information about the size of the study area of interest on the ground in m^2. Next we need to refine our radiometric measurement techniques to include information on what direction the radiant flux is leaving (exiting) the study area.

Radiance

Radiance is the most precise remote sensing radiometric measurement. *Radiance* (L_λ) is the radiant intensity per unit of projected source area in a specified direction. It is measured in watts per meter squared per steradian (W m^{-2} sr^{-1}). The concept of radiance can be understood by examining the components shown in Figure 6-19. First, the radiant flux leaves the projected

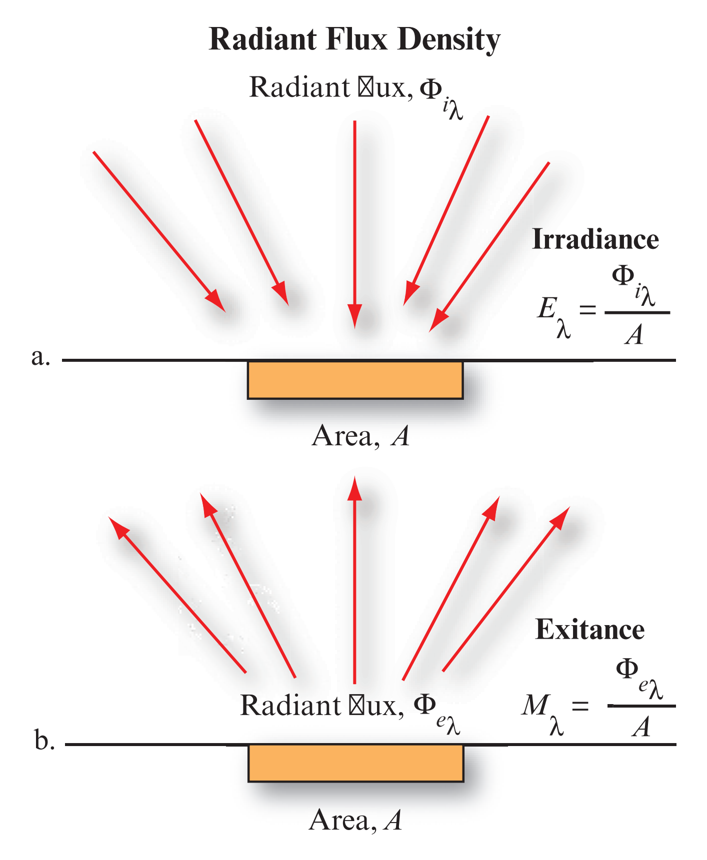


FIGURE 6–18 The concept of radiant flux density for an area on the surface of the Earth. a) *Irradiance* is a measure of the amount of radiant flux incident upon a surface per unit area of the surface measured in watts m^{-2}. b) *Exitance* is a measure of the amount of radiant flux leaving a surface per unit area of the surface measured in watts m^{-2}.

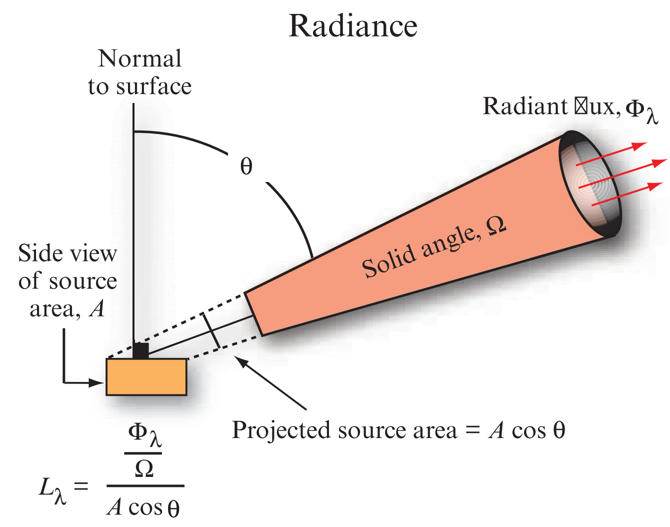


FIGURE 6–19 The concept of *radiance* leaving a specific projected source area on the ground, in a specific direction, and within a specific solid angle.

source area in a specific direction toward the remote sensor. We are not concerned with any other radiant flux that might be leaving the source area in any other direction. We are interested only in the radiant flux in certain wavelengths (Φ_λ) leaving the projected source area (A) within a certain direction ($\cos\ \theta$) and solid angle (Ω) (Milman, 1999):

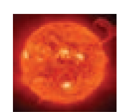
$$L_\lambda = \frac{\frac{\Phi_\lambda}{\Omega}}{A\cos\ \theta}. \quad (6.22)$$

One way of visualizing the solid angle is to consider what you would see if you were in an airplane looking through a telescope at the ground. Only the energy that exited the terrain and came up to and through the telescope in a specific solid angle (measured in steradians) would be intercepted by the telescope and viewed by your eye. Therefore, a solid angle is similar to a three-dimensional cone (or tube) that funnels radiant flux from a specific point source on the terrain toward the sensor system. Hopefully, energy from the atmosphere or other terrain features does not become scattered into the solid angle field of view and contaminate the radiant flux from the area of interest on the ground. Unfortunately, this is not often the case because scattering in the atmosphere and from other nearby areas on the ground can contribute spurious spectral energy, which enters into the solid angle field of view.



Energy–Matter Interactions in the Atmosphere Once Again

The radiant flux reflected or emitted from the Earth's surface once again enters the atmosphere, where it interacts with the various gases, water vapor, and particulates. Thus, atmospheric scattering, absorption, reflection, and refraction influence the radiant flux once again before the energy is recorded by the remote sensing system.



Energy–Matter Interactions at the Sensor System

Finally, energy–matter interactions take place when the energy reaches the remote sensing instrument. If an analog aerial camera is being used, then the radiance will interact with the camera filter, the optical glass lens, and finally the film emulsion with its very light-sensitive silver halide crystals. The emulsion must then be developed and printed before an analog hard copy is available for analysis. Rather than storing a latent

image on film, an optical–mechanical detector will digitally record the number of photons in very specific wavelength regions reaching the sensor.



Correcting Remote Sensing Detector Error

Ideally, the radiance recorded by a remote sensing system in various bands is an accurate representation of the radiance actually leaving the feature of interest (e.g., soil, vegetation, rocks, water, and/or urban land cover) on the Earth's surface. Unfortunately, noise (error) can enter the data-collection system at several points. For example, radiometric error in remotely sensed data may be introduced by the sensor system itself when the individual detectors do not function properly or are improperly calibrated. Several of the more common remote sensing system–induced radiometric errors are:

- random bad pixels (shot noise),
- line-start/stop problems,
- line or column drop-outs,
- partial line or column drop-outs, and
- line or column striping.

Sometimes digital image processing can recover the miscalibrated spectral information and make it relatively compatible with the correctly acquired data in the scene. Unfortunately, sometimes only cosmetic adjustments can be made to compensate for the fact that no data of value were acquired.

Random Bad Pixels (Shot Noise)

Sometimes an individual detector does not record spectral data for an individual pixel. When this occurs randomly, it is called a *bad pixel*. When there are numerous random bad pixels found within the scene, it is called *shot noise* because it appears that the image was shot by a shotgun. Normally these bad pixels contain values of 0 or 255 (in 8-bit data) in one or more of the bands. Shot noise is identified and repaired using the following methodology.

It is first necessary to locate each bad pixel in the band k dataset. A simple thresholding algorithm makes a pass through the dataset and flags any pixel ($BV_{i,j,k}$) having a brightness value of zero (assuming values of 0 represent shot noise and not a real land cover such as water). Once identified, it is then possible to evaluate the eight pixels surrounding the flagged pixel, as shown below:

	col_{j-1}	col_j	col_{j+1}
row_{i-1}	BV_1	BV_2	BV_3
row_i	BV_8	BV_{ijk}	BV_4
row_{i+1}	BV_7	BV_6	BV_5

The mean of the eight surrounding pixels is co mputed using Equation 6.23 and the value substituted for $BV_{i,j,k}$ in the corrected image:

$$BV_{i,j,k} = \text{Int}\left[\left(\sum_{i=1}^{8} BV_i\right)/8\right]. \quad (6.23)$$

This operation is performed for every bad (shot noise) pixel in the dataset.

For example, Landsat Thematic Mapper band 7 imagery (2.08 – 2.35 μm) of the Santee Delta is shown in Figure 6-20. It contains two pixels along a bad scan line with values of zero. The eight brightness values surrounding each bad pixel are enlarged and annotated in Figure 6-20b. The Landsat TM band 7 image after shot noise removal is shown in Figure 6-20c.

Line or Column Drop-Outs

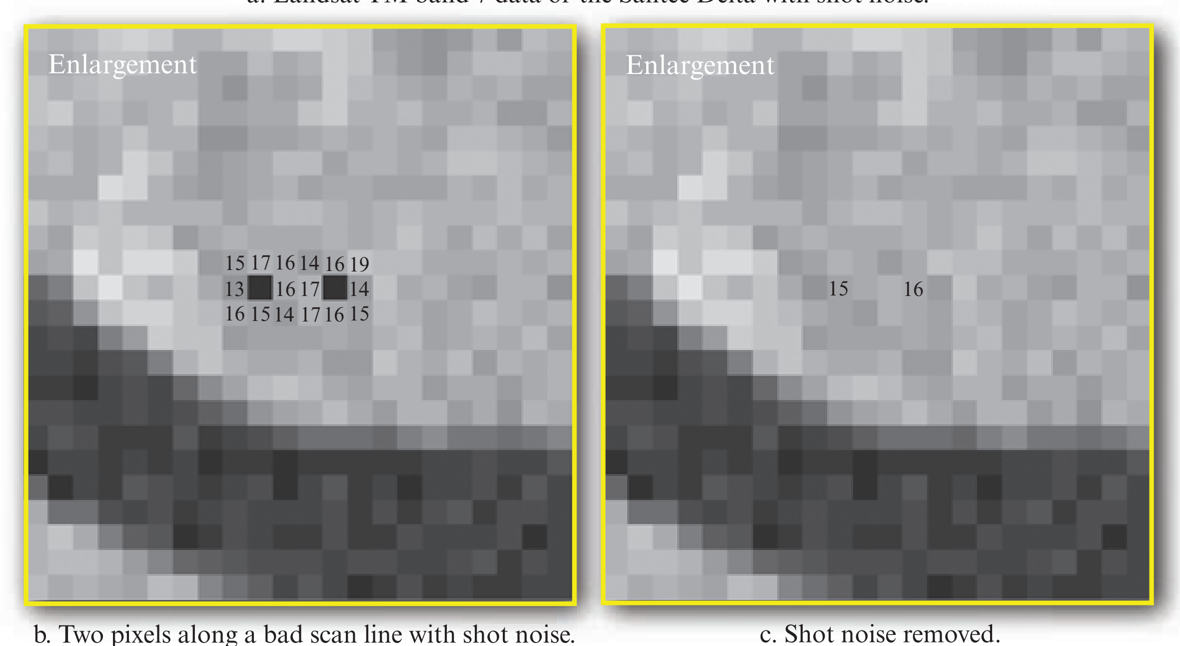
An entire line containing no spectral information may be produced if an individual detector in a scanning system (e.g., Landsat 7 ETM^+) fails to function properly. If a detector in a linear array (e.g., SPOT 5, Geo-Eye-1, WorldView-2) fails to function, this can result in an entire column of data with no spectral information. The bad line or column is commonly called a *line* or *column drop-out* and contains brightness values equal to zero. For example, if one of the 16 detectors in the Landsat Thematic Mapper sensor system fails to function during scanning, this can result in a brightness value of zero for every pixel, *j*, in a particular line, *i*. This *line drop-out* would appear as a completely black line in the band, *k*, of imagery. This is a serious condition because there is no way to restore data that were never acquired. However, it is possible to improve the visual interpretability of the data by introducing estimated brightness values for each bad scan line.

It is first necessary to locate each bad line in the image dataset. A simple thresholding algorithm makes a pass through the dataset and flags any scan line having a mean brightness value at or near zero. Once identified,

Correction of Shot Noise



a. Landsat TM band 7 data of the Santee Delta with shot noise.



b. Two pixels along a bad scan line with shot noise.

c. Shot noise removed.

FIGURE 6–20 a) Landsat Thematic Mapper band 7 (2.08 – 2.35 μm) image of the Santee Delta in South Carolina. One of the 16 detectors exhibits serious striping and the absence of brightness values at pixel locations along a scan line. b) An enlarged view of the bad pixels with the brightness values of the eight surrounding pixels annotated. c) The brightness values of the bad pixels after shot noise removal. This image was not destriped. Original imagery courtesy of NASA.

it is then possible to collect the value for a pixel in the preceding line ($BV_{i-1,j,k}$) and the value for a pixel in the succeeding line ($BV_{i+1,j,k}$) and assign the output pixel ($BV_{i,j,k}$) in the drop-out line the average of these two brightness values:

$$BV_{i,j,k} = \mathrm{Int}\left(\frac{BV_{i-1,j,k} + BV_{i+1,j,k}}{2}\right). \quad (6.24)$$

This is performed for every pixel in a bad scan line. The result is an image consisting of interpolated data every *n*th line that is more visually interpretable than one with horizontal black lines running systematically throughout the entire image. This same cosmetic digital image processing procedure can be applied to *column drop-outs* produced by a linear array remote sensing system.

Partial Line or Column Drop-Outs

Occasionally an individual detector will function perfectly along a scan line and then for some unknown reason it will not function properly for *n* columns. Then sometimes the detector functions properly again for the remainder of the scan line. The result is a portion of a scan line with no data. This is commonly referred to as a *partial line* or *partial column drop-out problem.* This is a serious condition. It usually cannot be dealt with systematically because it occurs randomly. If the portion of the image with the drop-out problem is particularly important, then the analyst must manually correct the problem pixel by pixel. The analyst must go into the dataset and compute the mean of the brightness values above and below each bad pixel and place the average in the bad pixel location. This is done for every pixel in the bad portion of the scan line.

Line-Start Problems

Occasionally, scanning systems fail to collect data at the beginning of a scan line, or they place the pixel data at inappropriate locations along the scan line. For example, all of the pixels in a scan line might be systematically shifted just one pixel to the right. This is called a *line-start* problem. Also, a detector may abruptly stop collecting data somewhere along a scan and produce results similar to the line or column drop-out previously discussed. Ideally, when data are not collected, the sensor system would be programmed to remember what was not collected and place any good data in their proper geometric locations along the scan. Unfortunately, this is not always the case. For example, the first pixel (column 1) in band *k* on line *i* (i.e., $BV_{i,1,k}$) might be improperly located at column 50 (i.e., $BV_{i,50,k}$). If the line-start problem is always associated with a horizontal bias of 50 columns, it can be corrected using a simple horizontal adjustment. However, if the amount of the line-start displacement is random, it is difficult to restore the data without extensive human interaction on a line-by-line basis. A considerable amount of MSS data collected by Landsats 2 and 3 exhibits line-start problems.

Figure 6-21a depicts predawn thermal infrared remote sensor data of the Savannah River obtained on March 31, 1981. It contains several line-start problems that are visually apparent. Seven lines with line-start problems (which contain good radiant energy information) were adjusted horizontally (to the left) just one pixel to be in their proper position (Figure 6-21b). Accurate linear measurements can now be made from the adjusted data. In addition, the image is now much more visually appealing.

N-Line Striping

Sometimes a detector does not fail completely, but simply goes out of radiometric adjustment. For example, a detector might record spectral measurements over a dark, deep body of water that are almost uniformly 20 brightness values greater than the other detectors for the same band. The result would be an image with systematic, noticeable lines that are brighter than adjacent lines. This is referred to as *n-line striping.* The maladjusted line contains valuable information, but should be corrected to have approximately the same radiometric scale as the data collected by the properly calibrated detectors associated with the same band.

To repair systematic *n*-line striping, it is first necessary to identify the miscalibrated scan lines in the scene. This is usually accomplished by computing a histogram of the values for each of the *n* detectors that collected data over the entire scene (ideally, this would take place over a homogeneous area, such as a body of water). If one detector's mean or median is significantly different from the others, it is probable that this detector is out of adjustment. Consequently, every line and pixel in the scene recorded by the maladjusted detector may require a *bias* (additive or subtractive) correction or a more severe *gain* (multiplicative) correction. This type of *n*-line striping correction a) adjusts all the bad scan lines so that they have approximately the same radiometric scale as the correctly collected data and b) improves the visual interpretability of the data. It looks better.

For example, Figure 6-22a presents a Landsat TM color composite (RGB = bands 4,3,2) of the Santee River delta in South Carolina obtained in 1984. The area consists primarily of salt marsh. The image exhibits serious striping every 16 lines. A pass was made through the dataset to identify the maladjusted lines. A gain and a bias were computed and applied to each affected line. Figure 6-22b depicts the same three band color composite after destriping. The striping is noticeably subdued, although close inspection reveals some residual striping. It is usually difficult to remove all traces of striping.

A 35-band hyperspectral 2×2 m spatial resolution dataset of the Mixed Waste Management Facility on the Savannah River Site in Aiken, SC, was acquired on

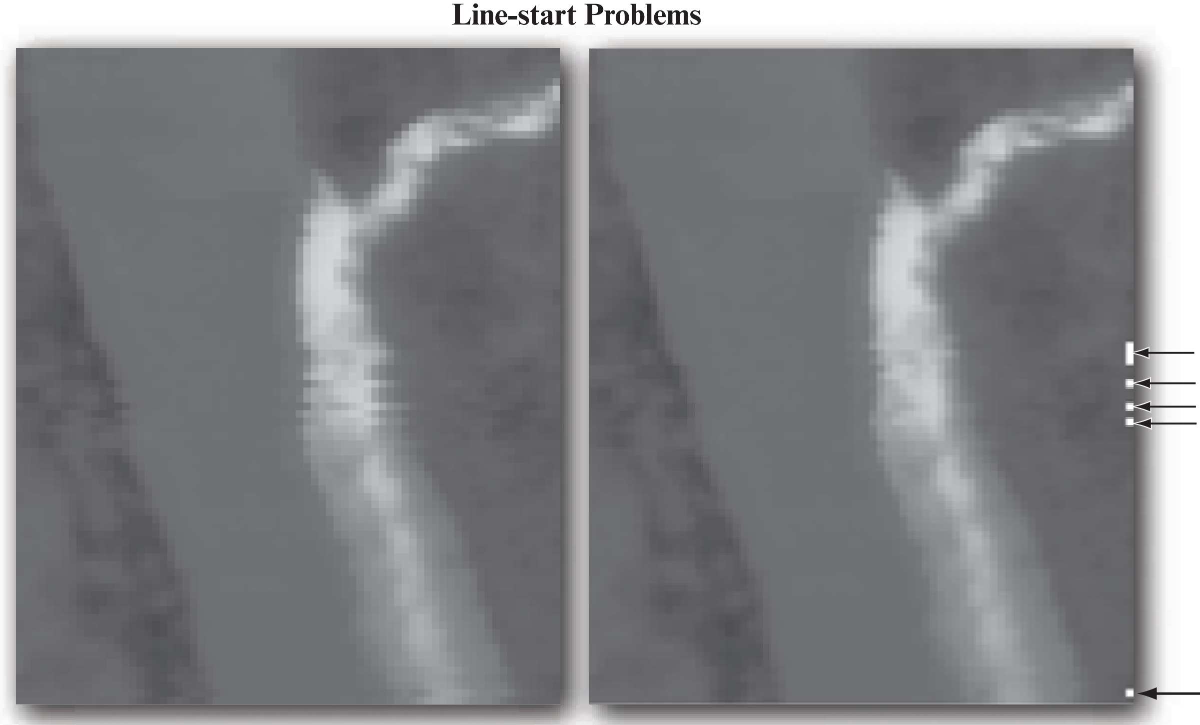


FIGURE 6–21 a) Infrared imagery of the Four Mile Creek thermal effluent plume entering the Savannah River on March 31, 1981. Seven lines with line-start problems are present. b) The result of translating each of the line-start problem lines one column to the left. The line-start problem was systematic in that offending lines only needed to be adjusted by one column. The line-start problem was unsystematic because not all of the lines in the dataset exhibited the line-start problem. Therefore, it was necessary to manually correct (using digital image processing) each of the line-start problem lines.

July 31, 2002 (Figure 6-23a). The level, clay-capped hazardous waste site fields are covered with Bahia grass or Centipede grass. A horizontal spectral profile through band 5 (red; centered at 633 nm) reveals good radiometric integrity along the columns of the dataset (Figure 6-24a). Unfortunately, a vertical spectral profile through the waste site exhibits a sawtooth pattern indicative of striping (Figure 6-24b). Using the previously described methodology, a gain and a bias were identified and applied to each offending fourth line in the dataset. The result is the radiometrically adjusted band 5 image shown in Figure 6-23cd. Examination of a vertical spectral profile through the corrected (destriped) dataset reveals that the sawtooth striping was greatly minimized but not completely removed (Figure 6-24c). The visual appearance of the band 5 data is noticeably improved.



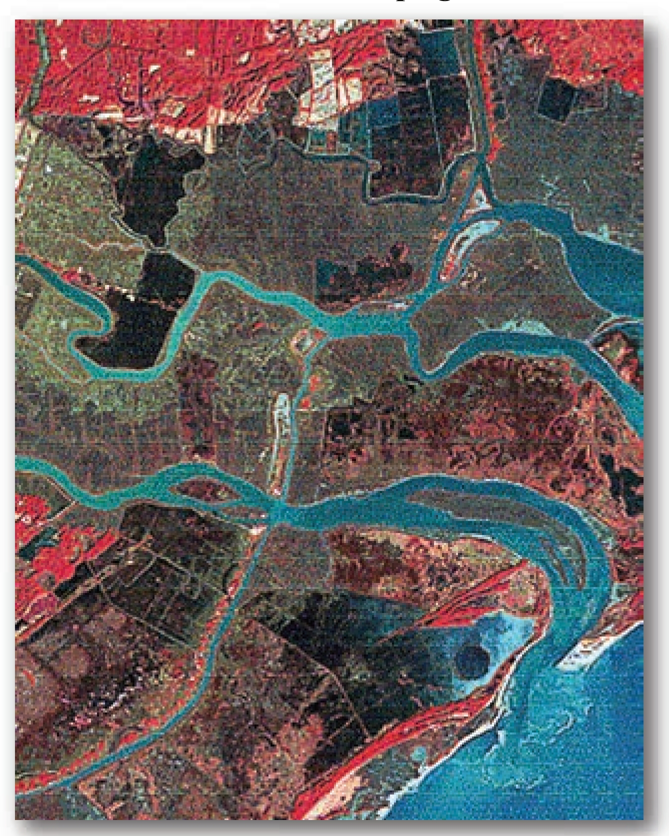
Remote Sensing Atmospheric Correction

Even when the remote sensing system is functioning properly, radiometric error may be introduced into the remote sensor data. The two most important sources of environmental attenuation are: 1) atmosphere attenuation caused by scattering and absorption in the atmosphere, and 2) topographic attenuation. However, it is important to first consider that it may not be necessary to atmospherically correct the remote sensor data for all applications. The decision to perform an atmospheric correction is a function of the nature of the problem, the type of remote sensing data available, the amount of *in situ* historical and/or concurrent atmospheric information available, and how accurate the biophysical information to be extracted from the remote sensing data must be (He and Chen, 2014).

Unnecessary Atmospheric Correction

Sometimes it is possible to ignore atmospheric effects in remote sensor data completely (Cracknell and Hayes, 1993; Song et al., 2001). For example, atmospheric correction is not always necessary for certain types of classification and change detection. Theoretical analysis and empirical results indicate that only when training data from one time or place must be extended through space and/or time is atmospheric correction necessary for image classification and many

Striping in Landsat Thematic Mapper Imagery



a. Landsat TM color composite of the Santee River delta in South Carolina.



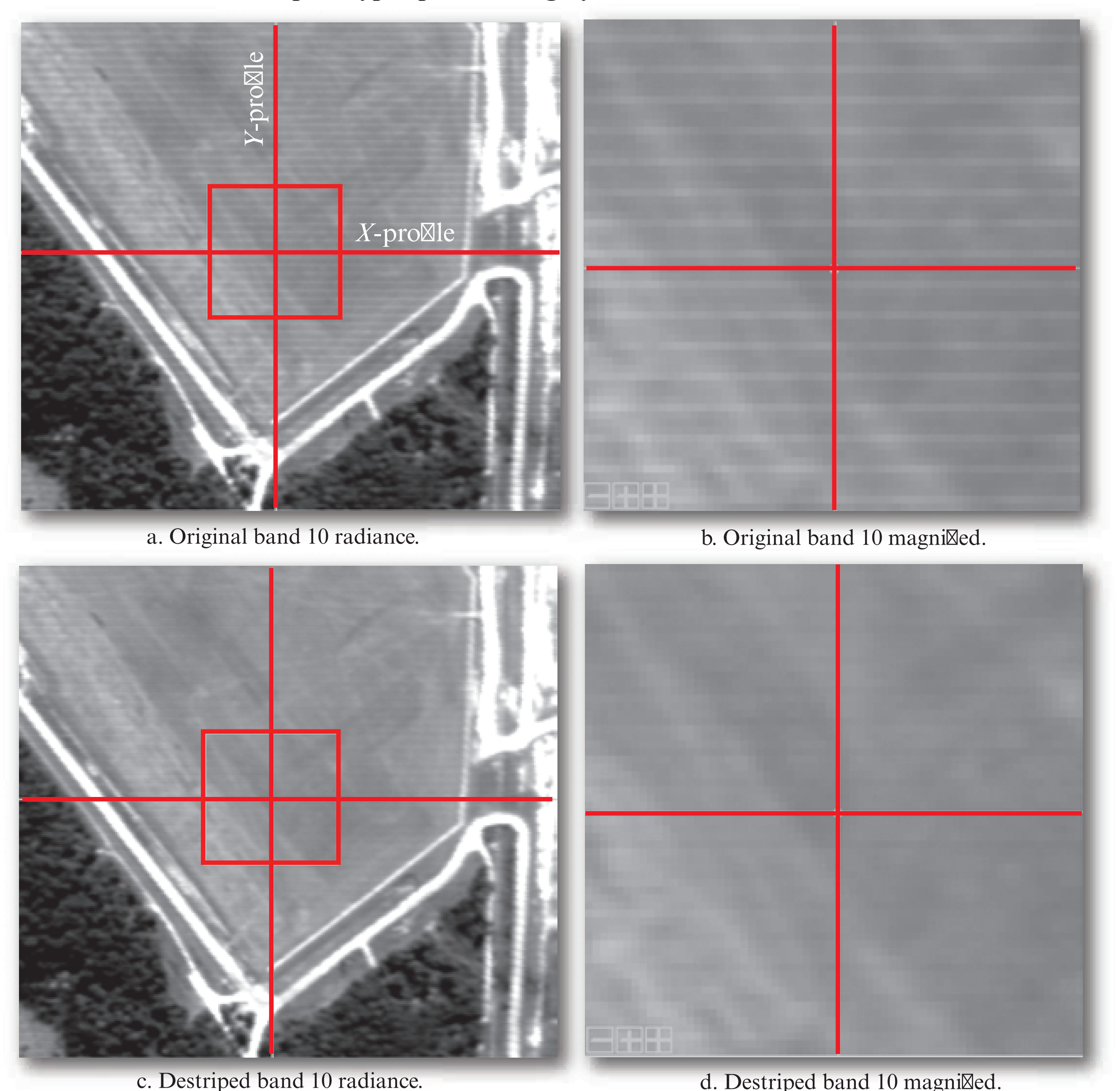
b. Landsat TM data destriped.

FIGURE 6–22 a) A Landsat Thematic Mapper image color composite (RGB = bands 4,3,2) of the Santee River delta in South Carolina in 1984. The area consists primarily of smooth cordgrass salt marsh (*Spartina alterniflora*). The image exhibits serious striping every 16 lines which have been highlighted by enhancing the contrast of the original TM dataset. A pass was made through the original dataset to identify the maladjusted lines. A gain and a bias were computed and applied to each affected line. b) The same study area is displayed after the data were destriped. Original imagery courtesy of NASA.

types of change detection (Song et al., 2001). For example, it is not generally necessary to perform atmospheric correction on a single date of remotely sensed data that will be classified using a maximum likelihood classification algorithm (discussed in Chapter 9). As long as the training data from the image to be classified have the same relative scale (corrected or uncorrected), atmospheric correction has little effect on classification accuracy (Song et al., 2001).

For example, consider land-cover classification using a single date of Landsat Thematic Mapper data. Rayleigh and other types of scattering normally add brightness to the visible bands (400 – 700 nm). Atmospheric absorption is the most important factor that reduces the brightness values of pixels in the near- and middle-infrared region (700 – 2,400 nm). Fortunately, the Landsat TM near- and middle-infrared bandwidths were carefully chosen to minimize the effects of atmospheric absorption. Therefore, if a single date of Landsat TM data is atmospherically corrected, it is likely that the primary effect will be a simple bias adjustment applied separately to each band. This action would adjust the minimum and maximum values of each band downward. Training class means extracted from the single-date image would change but the training class variance–covariance matrices should remain invariant. Therefore, the actual information content used in the maximum likelihood classification of the dataset would remain unchanged.

This logic also applies to certain types of change detection. For example, consider a change detection study involving two dates of imagery. If the two dates of imagery are analyzed independently using a standard classification algorithm and the resultant classification maps are compared using post-classification change detection logic (discussed in Chapter 12), there is no need to atmospherically correct the individual dates of remote sensor data. Similarly, atmospheric correction is not necessary when performing multidate composite image change detection (Jensen et al., 1993) where the



a. Original band 10 radiance.

b. Original band 10 magnified.

c. Destriped band 10 radiance.

d. Destriped band 10 magnified.

FIGURE 6–23 a) Original band 10 radiance (W m^{-2} sr^{-1}) data from a GER DAIS 3715 hyperspectral dataset of the Mixed Waste Management Facility on the Savannah River Site near Aiken, SC. The subset is focused on a clay-capped hazardous waste site covered with Bahia grass and Centipede grass. The 35-band data set was obtained at 2 × 2 m spatial resolution. The radiance values along the horizontal (X) and vertical (Y) profiles are summarized in Figure 6-24. b) An enlargement of the band 10 data. c) The band 10 data after destriping. d) An enlargement of the destriped data (Jensen et al., 2003).

change detection algorithm identifies change classes based on an analysis of all bands on both dates being placed in a single dataset (e.g., four TM bands from date 1 and four TM bands from date 2 are placed in a single eight-band dataset).

The general principle is that atmospheric correction is not necessary as long as the training data are *extracted* from the image (or composite image) under investigation and are not imported from another image obtained at another place or time.

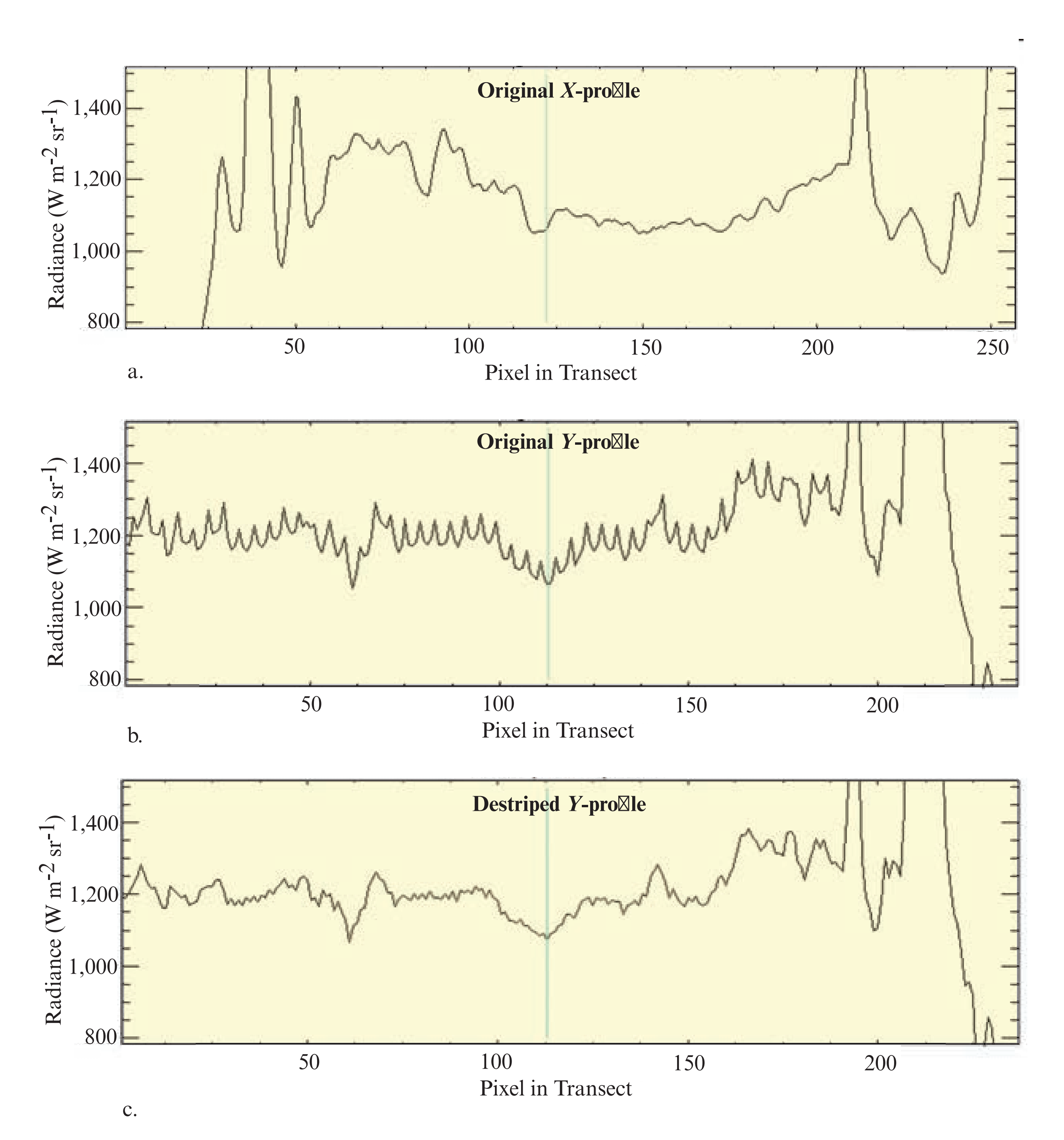


FIGURE 6–24 a) The radiance values along the horizontal (*X*) profile of the original band 10 radiance values in Figure 6-23. b) The radiance values along the vertical (*Y*) profile of the original band 10 radiance values in Figure 6-23. c) The radiance values along the vertical (*Y*) profile of the destriped band 10 radiance values. Note the reduction of the saw-toothed pattern in the destriped data (Jensen et al., 2003).

Necessary Atmospheric Correction

Sometimes it is essential that the remotely sensed data be atmospherically corrected. For example, it is usually necessary to atmospherically correct the remote sensor data if biophysical parameters are going to be extracted from water bodies (e.g., chlorophyll *a*, suspended sediment, temperature) or vegetation (e.g., biomass, leaf-area-index, chlorophyll, percent canopy closure). If the data are not corrected, the subtle differences in reflectance (or emittance) among the important constituents may be lost. Furthermore, if the biophysical measurements extracted from one image (e.g., biomass) are to be compared with the same biophysical information extracted from other images obtained on

different dates, then it is usually essential that the remote sensor data be atmospherically corrected.

For example, consider the case of the normalized difference vegetation index (NDVI) derived from Landsat Thematic Mapper (TM) band 3 (red) and band 4 (near-infrared) data:

$$\text{NDVI} = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}$$

which is used routinely to measure vegetation biomass and functional health in many decision-support systems such as the Africa Famine Early Warning System and Livestock Early Warning System (Jensen et al., 2002). Erroneous NDVI estimates can result in the loss of livestock and human life. Contributions from the atmosphere to NDVI are significant and can amount to 50% or more over thin or broken vegetation cover. Therefore, there is significant interest in removing the deleterious effects of the atmosphere in remotely sensed data that are used to compute NDVI estimates. The simple ratio (SR) vegetation index (TM4/TM3 for Landsat TM data) is also contaminated by the atmosphere (Song et al., 2001).

Great emphasis is being placed on the development of algorithms that can extract biophysical information from remotely sensed data for local, regional, and global applications. These data are then placed in various models and decision-support systems to monitor global processes and hopefully improve the quality of life. Consequently, great emphasis is being placed on the ability to derive measurements from the remotely sensed data that are accurate across vast distances and through time. **Signature extension** through space and time is becoming more important. The only way to extend signatures through space and time is to atmospherically correct each individual date of remotely sensed data.

Types of Atmospheric Correction

There are several ways to atmospherically correct remotely sensed data. Some are relatively straightforward while others are complex, being founded on physical principles and requiring a significant amount of information to function properly. This discussion will focus on two major types of atmospheric correction:

- *absolute atmospheric correction*, and
- *relative atmospheric correction*.

There are various methods that can be used to achieve absolute or relative atmospheric correction. The following sections identify the logic, some of the algorithms, and problems associated with each methodology.

1. It is possible to use a *model atmosphere* to correct the remotely sensed data. An assumed atmosphere is calculated based on the time of year, altitude, latitude, and longitude of the study area. This approach may be successful when atmospheric attenuation is relatively small compared with the signal from the terrain being remotely sensed (Cracknell and Hayes, 1993).
2. The use of a model atmosphere in conjunction with *in situ* atmospheric measurements acquired at the time of remote sensor data acquisition is even better. Sometimes, the *in situ* data can be provided by atmospheric sounding instruments found onboard the sensor platform. The atmospheric model may then be fine-tuned using the local condition information. This is referred to as *absolute radiometric correction,* and an example is provided later in the chapter.
3. Minimization of atmospheric attenuation is sometimes possible using multiple looks at the same object from different vantage points (e.g., fore and aft) or by looking at the same object using multiple bands of the spectrum. The goal is to try to have the information from the multiple looks or multiple bands cancel out the atmospheric effects. The multiple-look method suffers from the fact that the atmospheric paths for the multiple looks (e.g., fore and aft) may not be the same. Theoretically, the band-cancellation method should be capable of providing good results because it is using identical atmospheric paths for the channels that are being compared. This is called *relative radiometric correction,* and an image normalization example will be provided.

Absolute Radiometric Correction of Atmospheric Attenuation

Solar radiation (E_o) is largely unaffected as it travels through the vacuum of space (Figure 6-25). When it interacts with the Earth's atmosphere, however, it is selectively scattered and absorbed. The sum of these two forms of energy loss is called **atmospheric attenuation**. Serious atmospheric attenuation may 1) make it difficult to relate hand-held *in situ* spectroradiometer measurements of materials with remote measurements of these same materials, 2) make it difficult to extend spectral signatures through space and time, and 3) have an impact on classification accuracy within a scene if atmospheric attenuation varies significantly throughout the image (Cracknell and Hayes, 1993; Kaufman et al., 1997).

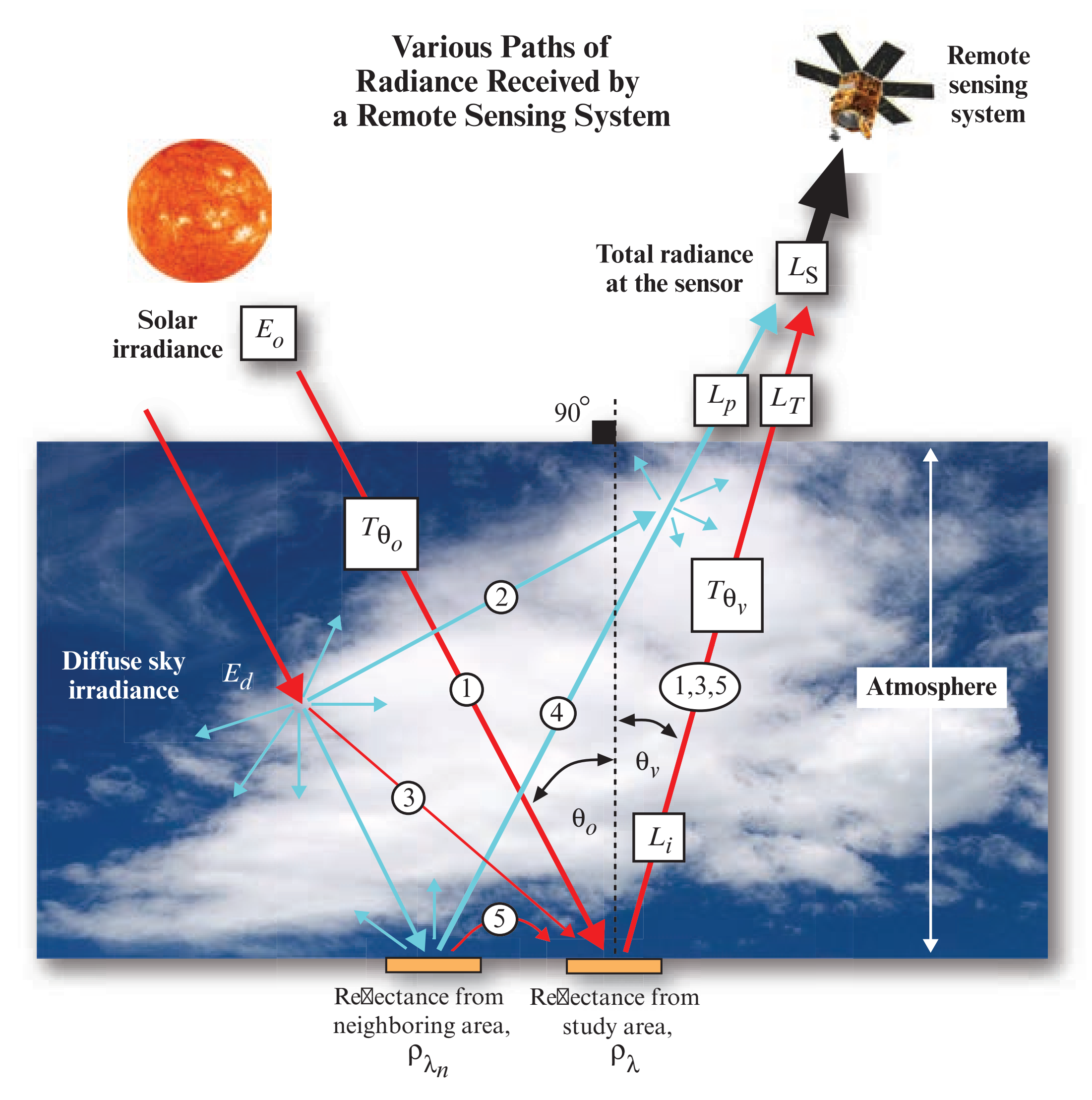


FIGURE 6–25 *Radiance* (L_T) from paths 1, 3, and 5 contains intrinsic valuable spectral information about the target of interest. Conversely, the *path radiance* (L_p) from paths 2 and 4 includes diffuse sky irradiance or radiance from neighboring areas on the ground. This *path radiance* generally introduces unwanted radiometric noise in the remotely sensed data and complicates the image interpretation process.

The general goal of **absolute radiometric correction** is to turn the digital brightness values recorded by a remote sensing system into **scaled surface reflectance** values (Du et al., 2002). These values can then be compared or used in conjunction with scaled surface reflectance values obtained anywhere else on the planet.

A considerable amount of research has been carried out to address the problem of correcting images for atmospheric effects. These efforts have resulted in a number of atmospheric radiative transfer codes (models) that can provide realistic estimates of the effects of atmospheric scattering and absorption on imagery (e.g., Kruse, 2004; Alder-Golden et al., 2005; Gao et al., 2009; Agrawal and Sarup, 2011; Richter and Schlapfer, 2014). Once these effects have been identified for a specific date of imagery, each band and/or pixel in the scene can be adjusted to remove the effects of scattering and/or absorption. The image is then considered to be atmospherically corrected.

Unfortunately, the application of these codes to a specific scene and date also requires knowledge of both the sensor spectral profile and the atmospheric properties at the same time. Atmospheric properties are difficult to acquire even when planned. For most historic satellite data, they are not available. Even today, accurate scaled surface reflectance retrieval is not operational for the majority of satellite image sources used for land-cover change detection. One exception is NASA's Moderate Resolution Imaging Spectroradiometer (MODIS), for which surface reflectance products are available. Nevertheless, we will proceed with a very general discussion of the important issues associated

with absolute atmospheric correction and then provide examples of how absolute radiometric correction is performed.

Target and Path Radiance

Ideally, the radiance (L) recorded by the camera or detector is a true function of the amount of radiance leaving the target terrain within the instantaneous-field-of-view (IFOV) at a specific solid angle, as previously discussed. Unfortunately, other radiant energy may enter into the field of view from various other paths and introduce confounding noise into the remote sensing process. Therefore, additional radiometric variable definitions are needed to identify the major sources and paths of this energy (Green, 2003). The variables are summarized in Table 6-5. The various paths and factors that determine the radiance reaching the remote sensing detector(s) are summarized in Figure 6-25, including:

- *Path 1* contains the spectral solar irradiance (E_{o_λ}) that was attenuated very little before illuminating the terrain within the IFOV. Notice that we are interested in the solar irradiance from a specific solar zenith angle (θ_o) and that the amount of irradiance reaching the terrain is a function of the atmospheric transmittance at this angle (T_{θ_o}). If all of the irradiance makes it to the ground, then the atmospheric transmittance (T_{θ_o}) equals one. If none of the irradiance makes it to the ground, then the atmospheric transmittance is zero.
- *Path 2* contains spectral diffuse sky irradiance (E_{d_λ}) that never even reaches the Earth's surface (the target study area) because of scattering in the atmosphere. Unfortunately, such energy is often scattered directly into the IFOV of the sensor system. As previously discussed, Rayleigh scattering of blue light contributes much to this diffuse sky irradiance. That is why the blue band image produced by a remote sensor system is often much brighter than any of the other bands. It contains much unwanted diffuse sky irradiance that was inadvertently scattered into the IFOV of the sensor system. Therefore, if possible, we want to minimize its effects. Green (2003) refers to the quantity as the upward reflectance of the atmosphere (E_{du_λ}).
- *Path 3* contains energy from the Sun that has undergone some Rayleigh, Mie, and/or nonselective scattering and perhaps some absorption and re-emission before illuminating the study area. Thus, its spectral composition and polarization may be somewhat different from the energy that reaches the ground from path 1. Green (2003) refers to this quantity as the downward reflectance of the atmosphere (E_{dd_λ}).

TABLE 6–5 Radiometric variables used in remote sensing.

Radiometric Variables	
E_o	= solar irradiance at the top of the atmosphere (W m^{-2})
E_{o_λ}	= spectral solar irradiance at the top of the atmosphere (W m^{-2} μm^{-1})
E_d	= diffuse sky irradiance (W m^{-2})
E_{d_λ}	= spectral diffuse sky irradiance (W m^{-2} μm^{-1})
E_{du_λ}	= the *upward* reflectance of the atmosphere
E_{dd_λ}	= the *downward* reflectance of the atmosphere
E_g	= global irradiance incident on the surface (W m^{-2})
E_{g_λ}	= spectral global irradiance on the surface (W m^{-2} μm^{-1})
τ	= normal atmospheric optical thickness
T_θ	= atmospheric transmittance at an angle θ to the zenith
θ_o	= solar zenith angle
θ_v	= view angle of the satellite sensor (or scan angle)
μ	= cos θ
ρ_λ	= surface target reflectance at a specific wavelength
ρ_{λ_n}	= reflectance from a neighboring area
L_s	= total radiance at the sensor (W m^{-2} sr^{-1})
L_t	= total radiance from the target of interest toward the sensor (W m^{-2} sr^{-1})
L_i	= intrinsic radiance of the target (W m^{-2} sr^{-1}) (i.e., what a hand-held radiometer would record on the ground without intervening atmosphere)
L_p	= path radiance from multiple scattering (W m^{-2} sr^{-1})

- *Path 4* contains radiation that was reflected or scattered by nearby terrain (ρ_{λ_n}) covered by snow, concrete, soil, water, and/or vegetation into the IFOV of the sensor system. The energy does not actually illuminate the study area of interest. Therefore, if possible, we would like to minimize its effects.
- *Path 5* is energy that was also reflected from nearby terrain into the atmosphere, but then scattered or reflected onto the study area.

Therefore, for a given spectral interval in the electromagnetic spectrum (e.g., λ_1 to λ_2 could be 0.6 – 0.7 μm or red light), the total solar irradiance reaching the *Earth's surface,* E_{g_λ}, is an integration of several important components:

$$E_{g_\lambda} = \int_{\lambda_1}^{\lambda_2} (E_{o_\lambda} T_{\theta_o} \cos\theta_o + E_{d_\lambda}) d\lambda \text{ (W m}^{-2} \mu\text{m}^{-1}). \quad (6.25)$$

It is a function of the spectral solar irradiance at the top of the atmosphere (E_{o_λ}) multiplied by the atmospheric transmittance (T_{θ_o}) at a certain solar zenith angle (θ_o) plus the contribution of spectral diffuse sky irradiance (E_{d_λ}).

Only a small amount of this irradiance is actually reflected by the terrain in the direction of the satellite sensor system. If we assume the surface of Earth is a diffuse reflector (a Lambertian surface), the total amount of radiance exiting the target study area (L_T) toward the sensor is:

$$L_T = \frac{1}{\pi} \int_{\lambda_1}^{\lambda_2} \rho_\lambda \, T_{\theta_v} (E_{o_\lambda} T_{\theta_o} \cos\theta_o + E_{d_\lambda}) d\lambda . \quad (6.26)$$

The average surface target reflectance (ρ_λ) is included because the vegetation, soil, and water within the IFOV selectively absorb some of the incident energy. Therefore, not all of the energy incident to the IFOV (E_{g_λ}) leaves the IFOV. In effect, the terrain acts like a filter, selectively absorbing certain wavelengths of light while reflecting others. Note that the energy exiting the terrain is at an angle (θ_v), requiring the use of an atmospheric transmittance factor T_{θ_v} once again.

It would be wonderful if the total radiance recorded by the sensor, L_S, equaled the radiance returned from the target study area of interest, L_T. Unfortunately, $L_S \neq L_T$ because there is some additional radiance from different *paths* that may fall within the IFOV of the sensor system detector (Figure 6-25). This is often called *path radiance*, L_P. Thus, the total radiance recorded by the sensor becomes:

$$L_S = L_T + L_P \quad (\text{W m}^{-2} \text{ sr}^{-1}). \quad (6.27)$$

We see from Equation 6.27 and Figure 6-25 that the path radiance (L_P) is an intrusive (bad) component of the total amount of radiance recorded by the sensor system (L_S). It is composed of radiant energy primarily from the diffuse sky irradiance (E_d) from path 2 as well as the reflectance from nearby ground areas ρ_{λ_n} from path 4. Path radiance introduces error to the remote sensing data-collection process. It can impede our ability to obtain accurate spectral measurements.

A great deal of research has gone into developing methods to remove the contribution of path radiance (L_P). Methods for computing path radiance are summarized in Richards (2013). Radiative transfer model programs such as MODTRAN, Second Simulation of the Satellite Signal in the Solar Spectrum (6S) and others may be used to predict path radiance on a particular day for a particular study area (e.g., Alder-Golden et al., 1999, 2005; Matthew et al., 2000; Gao et al., 2009; Richter and Schlapfer, 2014). Such information can then be used to remove the path radiance (L_P) contribution to the remote sensing signal (L_S).

Atmospheric Transmittance

To understand how to remove atmospheric attenuation, it is useful to review the fundamental mechanisms of atmospheric scattering and absorption related to atmospheric transmittance. In the absence of atmosphere, the transmittance of solar radiant energy to the ground would be 100%. However, because of absorption and scattering, not all of the radiant energy reaches the ground. The amount that does reach the ground, relative to that for no atmosphere, is called *transmittance.* Atmospheric transmittance (T_θ) may be computed as:

$$T_\theta = e^{-\tau/(\cos\theta)} \quad (6.28)$$

where τ is the normal atmospheric optical thickness and θ can represent either θ_o or θ_v in Figure 6-25 (i.e., the ability of the atmosphere to transmit radiant flux from the Sun to the target, T_{θ_o}, or the ability of the atmosphere to transmit radiant flux from the target to the sensor system, T_{θ_v}). The optical thickness of the atmosphere at certain wavelengths, $\tau(\lambda)$, equals the sum of all the attenuating coefficients, which are made up primarily of Rayleigh scattering, τ_m, Mie scattering, τ_p, and selective atmospheric absorption, τ_a:

$$\tau(\lambda) = \tau_m + \tau_p + \tau_a \quad (6.29)$$

where $\tau_a = \tau_{H_2O} + \tau_{O_2} + \tau_{O_3} + \tau_{CO_2}$.

As previously discussed, Rayleigh scattering of gas molecules occurs when the diameter of the gas molecule (d) is less than the size of the incident wavelength ($d < \lambda$) and is inversely proportional to the fourth power of the wavelength, $1/\lambda^4$. Aerosol (Mie) scattering occurs when $d = \lambda$ and is primarily a function of water vapor, dust, and other aerosol particles in the atmosphere. Selective absorption of radiant energy in the atmosphere is wavelength-dependent. Most of the optical region from 0.4 to 1.0 μm is dominated by water and ozone absorption attenuation. Atmospheric absorption by water vapor and other gases in the atmosphere mostly affects radiation of wavelengths > 0.8 μm. Thus, atmospheric scattering may add brightness to, whereas atmospheric

absorption may subtract brightness from, landscape spectral measurements.

Diffuse Sky Irradiance

Path 1 in Figure 6-25 contains radiation that followed a direct path from the Sun, to the target study area within the IFOV, to the remote sensing detector. Unfortunately, some additional scene irradiance comes from scattered skylight. Path 3 in Figure 6-25 contains radiation that has undergone some scattering before illuminating the study area. Thus, its spectral composition may be somewhat different. Similarly, path 5 contains energy that was reflected from a nearby area on the ground into the atmosphere and then scattered once again onto the study area. The total diffuse sky irradiance at the pixel is E_d.

It is now possible to determine how these atmospheric effects (transmittance, diffuse sky irradiance, and path radiance) affect the radiance measured by the remote sensing system. First, however, because the bandwidths used in remote sensing are relatively narrow (e.g., 0.5 to 0.6 μm), it is possible to restate Equations 6.25 and 6.26 without the integral. For example, the total irradiance at the Earth's surface may be written:

$$E_g = E_{o\Delta\lambda} T_{\theta_o} \cos\theta_o \Delta\lambda + E_d, \quad (6.30)$$

where $E_{o\Delta\lambda}$ is the average spectral irradiance in the band interval $\Delta\lambda = \lambda_2 - \lambda_1$. The total radiance transmitted through the atmosphere toward the sensor (L_T) becomes:

$$L_T = \frac{1}{\pi}\rho T_{\theta_v}(E_{o\Delta\lambda} T_{\theta_o} \cos\theta_o \Delta\lambda + E_d). \quad (6.31)$$

The total radiance reaching the sensor then becomes:

$$L_S = \frac{1}{\pi}\rho T_{\theta_v}(E_{o\Delta\lambda} T_{\theta_o} \cos\theta_o \Delta\lambda + E_d) + L_P, \quad (6.32)$$

which may be used to relate brightness values in remotely sensed data to measured radiance, using the equation:

$$L_S = (K \times BV_{i,j,k}) + L_{min}, \quad (6.33)$$

where

K = radiance per bit of sensor count rate = $(L_{max} - L_{min})/C_{max}$
BV_{ijk} = brightness value of a pixel
C_{max} = maximum value in dataset (e.g., 8-bit = 255)
L_{max} = radiance measured at detector saturation ($W\ m^{-2}\ sr^{-1}$)
L_{min} = lowest radiance measured by a detector ($W\ m^{-2}\ sr^{-1}$)

The L_{min} and L_{max} values are available for all the major remote sensing systems.

Atmospheric Correction Based on Radiative Transfer Modeling

Most current radiative transfer-based atmospheric correction algorithms can compute much of this same information if a) the user provides fundamental atmospheric characteristic information to the program, or b) certain atmospheric absorption bands are present in the remote sensing dataset. For example, most radiative transfer-based atmospheric correction algorithms require that the user provide:

- latitude and longitude of the remotely sensed image scene,
- the date and precise time that the remote sensing data were collected,
- image acquisition altitude (e.g., 20 km AGL),
- mean elevation of the scene (e.g., 200 m ASL),
- an atmospheric model (e.g., mid-latitude summer, mid-latitude winter, tropical),
- radiometrically calibrated image radiance data (i.e., the data *must* be in the form $W\ m^2\ \mu m^{-1}\ sr^{-1}$),
- information about each specific band (i.e., its mean and full-width at half-maximum (FWHM), and
- local atmospheric visibility at the time of remote sensing data collection (e.g., 10 km, obtained from a nearby airport if possible).

These parameters are then input to the atmospheric model selected (e.g., mid-latitude summer) and used to compute the absorption and scattering characteristics of the atmosphere at the instance of remote sensing data collection. These atmospheric characteristics are then used to invert the remote sensing radiance to **scaled surface reflectance**. Many of these atmospheric correction programs derive the scattering and absorption information they require from robust atmosphere radiative transfer code such as MODTRAN 4+ (Alder-Golden et al., 1999; 2005) or Second Simulation of the Satellite Signal in the Solar Spectrum (6S) (Vermote et al., 1997, 2002; Gao et al., 2009).

Numerous atmospheric correction algorithms have been developed based on the use of radiative transfer principles (Gao et al., 2009). Several of the most important are ACORN, QUAC, FLAASH, and ATCOR.

It is instructive briefly to review some of their characteristics.

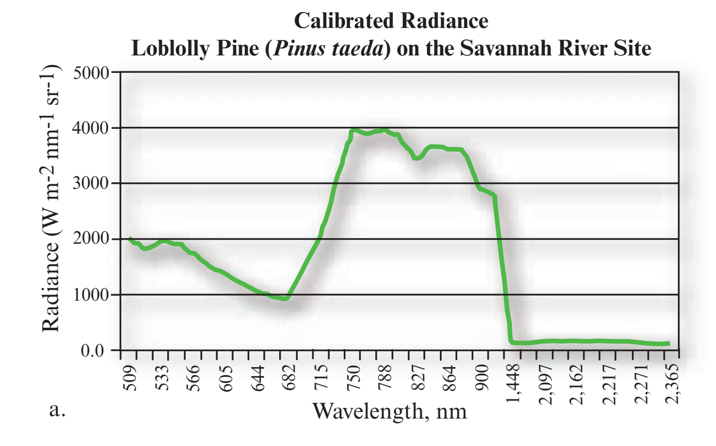
- **ACORN**—**A**tmospheric **COR**rection **N**ow is based on MODTRAN 4 radiative transfer code (San and Suzen, 2010; ImSpec ACORN, 2014). It performs atmospheric correction of both multispectral and hyperspectral imagery in the region from 250 – 2,500 nm. It is designed to work with all airborne and spaceborne calibrated remote sensing systems such as Hyperion, ASTER, Landsat ETM$^+$, AVIRIS, SPOT, GeoEye-1, etc. It defines the relationship from contributions of the exo-atmospheric solar source, a homogenous plane parallel atmosphere, and the surface with respect to the radiance measured by an Earth-looking remote sensor (ImSpec ACORN, 2014). Kruse (2004) suggests that all model-based atmospheric correction algorithms such as ACORN and FLAASH basically use the radiative transfer model shown below, though each model uses a slightly different version:

$$L_S = [E_{o_\lambda} T_\lambda \rho_\lambda \cos(\theta)] + L_P, \tag{6.34}$$

where L_S is the total radiance measured by the sensor, E_{o_λ} is the solar spectral irradiance at the top of the atmosphere, T_λ is the total atmospheric transmittance, ρ_λ is the spectral reflectance of the surface, and L_P is the path radiance.

Current atmospheric correction programs assume that the Earth's surface is horizontal and has Lambertian reflectance. This is because for real data we typically don't have enough information to make the topographic correction. The end result is called *scaled surface reflectance* or *apparent reflectance*. The scaled surface reflectance can be converted to surface reflectance if the surface topography is known (Kruse, 2004). The principal outputs of ACORN are a water vapor image and scaled surface reflectance data.

For example, consider the hyperspectral data of the Savannah River Site that was atmospherically corrected and transformed from radiance to scaled surface reflectance using ACORN. Figure 6-26a depicts the original radiance (in W m^{-2} nm^{-1} sr^{-1}) of a pixel of loblolly pine (*Pinus taeda*) in 35 bands of GER DAIS 3715 hyperspectral imagery from 509 nm to 2,365 nm. Figure 6-26b is the scaled surface reflectance of the same pixel. The Loblolly pine pixel now appears as it should, with a slight peak of reflectance in the green at approximately 566 nm, chlorophyll absorption in the red centered at approximately 682 nm, and increased near-infrared reflectance (Jensen et al., 2003).





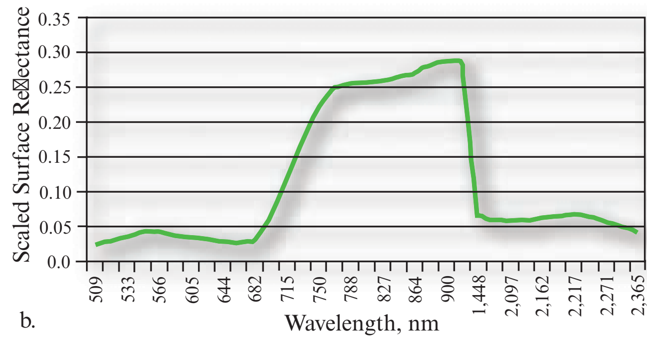


FIGURE 6–26 a) Calibrated radiance data of a Loblolly Pine pixel on the Savannah River Site recorded by the GER DAIS 3715 hyperspectral remote sensing system. The 35 bands range from 509 – 2,365 nm. b.) The same pixel transformed into scaled surface reflectance using the radiative transfer Atmospheric CORrection Now (ACORN) algorithm (based on Jensen et al., 2003).

- **FLAASH**—ENVI's **F**ast **L**ine-of-sight **A**tmospheric **A**nalysis of **S**pectral **H**ypercubes was developed by the Air Force Phillips Laboratory, Hanscom AFB and Spectral Sciences, Inc. FLAASH is an atmospheric correction program that corrects remote sensor data in the region from 400 – 3,000 nm (Exelis QUAC and FLAASH, 2009; San and Suzen, 2010; Exelis Atmospheric Correction, 2014). FLAASH can be used to atmospherically correct data from hyperspectral sensors such as HyMAP, AVIRIS, HYDICE, HYPERION, Probe-1, CASI, and AISA and multispectral sensors such as ASTER, IRS, Landsat, RapidEye, and SPOT.

  FLAASH uses MODTRAN simulations of spectral radiance computed for various atmospheric, water vapor, and viewing conditions (solar angles) over a range of surface reflectances to establish lookup tables for the atmospheric parameters of column water vapor, aerosol type, and visibility for subsequent use (Kruse, 2004; Alder-Golden et al., 2005). The 1.13 μm

AVIRIS Image of a Part of Jasper Ridge Biological Preserve in San Mateo County, CA, collected on April 3, 1998



a. Natural color composite consisting of red (band 28 centered on 636 nm), green (band 19 centered on 547 nm), and blue (band 10 centered on 458 nm).

Spectral Profile of Uncorrected and FLAASH® Atmospherically Corrected Pixel

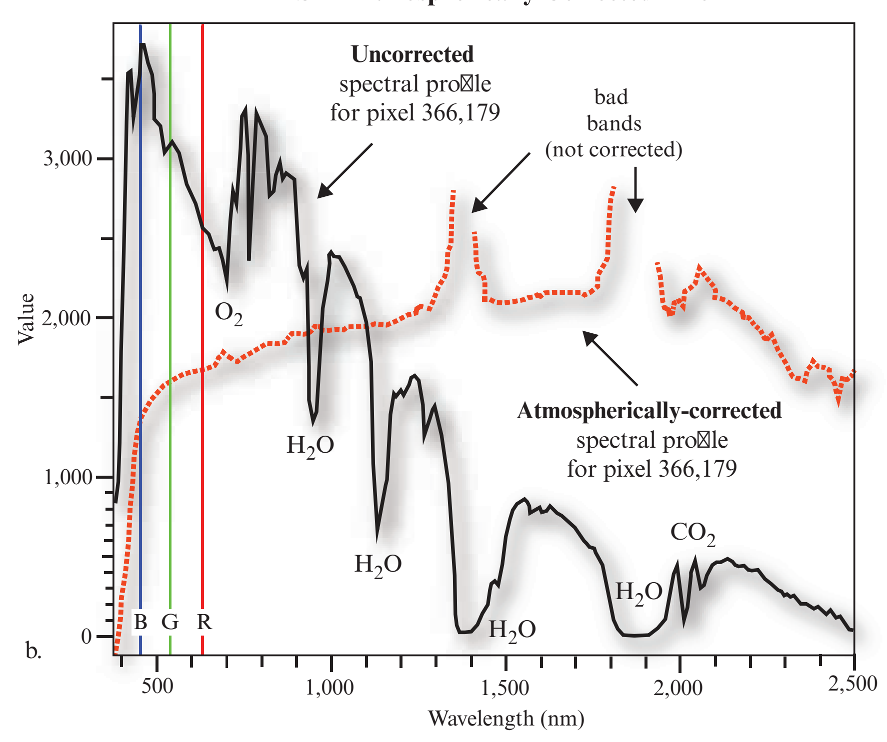


FIGURE 6-27 a) AVIRIS 20 × 20 m hyperspectral image consisting of 224 bands (AVIRIS data courtesy of NASA Jet Propulsion Lab). b) Uncorrected spectral profile of a pixel of concrete at column 369 and row 179 (value scale = radiance). ENVI FLAASH atmospherically corrected spectral profile of the same pixel (value scale = scaled percent reflectance × 1,000).

water band is used to estimate water vapor, and a ratio of in-band and out-of-band radiance values allows estimation of absorption band depths for a range of water vapor column densities. FLAASH also derives pressure altitudes by applying the same method to the oxygen 0.762 μm absorption band. The radiance spectra are extracted from the hyperspectral data and compared against the MODTRAN lookup tables on a pixel-by-pixel basis to determine scaled surface reflectance. FLAASH offers the additional option of correcting for light scattered from adjacent pixels (Alder-Golden et al., 1999; Matthew et al., 2003). It also produces a column water vapor image, a cloud map, and a visibility value for the scene (Felde et al., 2003; Kruse, 2004; Exelis FLAASH Multispectral, 2014).

An example of the application of FLAASH atmospheric correction to AVIRIS hyperspectral data is shown in Figure 6-27. The 20 × 20 m AVIRIS data of a part of the Jasper Ridge Biological Preserve consists of 224 bands and was acquired on April 3, 1998. A pixel centered on the concrete freeway at 366,179 is highlighted in Figure 6-27a. The uncorrected spectral profile of this pixel is shown in Figure 6-27b. Note the impact of various atmospheric constituents such as oxygen, water vapor, and carbon dioxide on the at-sensor radiance spectral profile. The FLAASH atmospherically-corrected pixel exhibits normal scaled percent reflectance characteristics (Figure 6-27b) with high reflectance throughout the blue, green, red, and near-infrared portion of the spectrum as expected (e.g., refer to the hand-held spectroradiometer reflectance measurement of concrete shown in Figure 6-17).

- **QUAC**—ENVI's **Qu**ick **A**tmospheric **C**orrection (QUAC) is a visible-near infrared through shortwave infrared (VNIR-SWIR) atmospheric correction method that can be used for both multispectral and hyperspectral imagery (Figure 6-28) (Bernstein et al., 2008; Exelis QUAC, 2014; Exelis Atmospheric Correction, 2014). Unlike ACORN and FLAASH, it determines atmospheric compensation parameters from the information contained within the image such as pixel endmember spectra (i.e., pure, unmixed pixels in the scene), without ancillary information (Figure 6-28). QUAC performs a more approximate atmospheric correction than FLAASH or other physics-based first-principles methods, generally producing reflectance spectra within approximately – 15% of the physics-based approaches. QUAC is based on the empirical finding that the average reflectance of a collection of diverse material spectra, such as the endmember spectra in a scene, is essentially scene-independent. QUAC also allows for any view or solar elevation angle. Should a sensor not have proper radiometric or wavelength calibration, or the solar illumination intensity be unknown (such as when a cloud deck is present), this approach still allows the retrieval of reasonably accurate reflectance spectra. QUAC provides (Exelis QUAC, 2014):

  - atmospheric correction of multispectral and hyperspectral data in the region from 400 – 2,500 nm.

  - atmospheric correction of the following types of remote sensor data: AISA, ASAS, AVIRIS, CAP ARCHER, COMPASS, HYCAS, HYDICE, HyMap, Hyperion, IKONOS, Landsat Thematic Mapper,

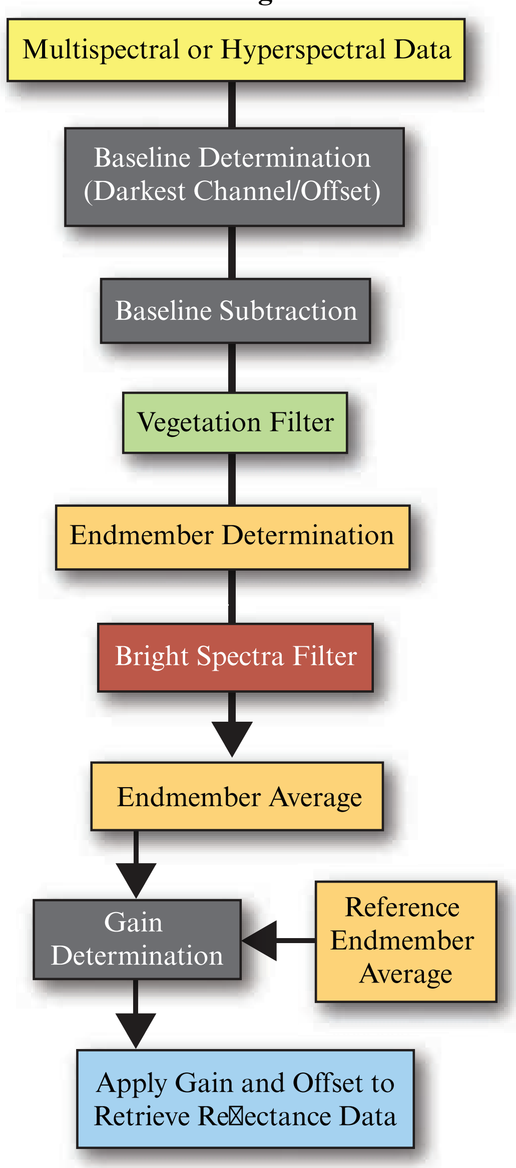


FIGURE 6–28 Quick Atmospheric Correction (QUAC) processing flow (adapted from Exelis QUAC and FLAASH, 2014.).

LASH, MASTER, MODIS, MTI, QuickBird, RGB, and unknown sensor types.

- an image of retrieved surface reflectance.

QUAC performs a fast and fairly accurate atmospheric correction if a) there are at least 10 diverse materials in a scene, and b) there are sufficiently dark pixels in a scene to allow for a good estimation of the baseline spectrum (Bernstein et al., 2008; Exelis QUAC, 2014; Exelis Atmospheric Correction, 2014). Agrawal and Sarup (2011) compared the results of applying both QUAC and FLAASH atmospheric corrections to EO-1 Hyperion hyperspectral data.

- **ATCOR**—The **AT**mospheric **COR**rection program was originally developed at DLR, the German Aerospace Centre (Richter and Schlapfer, 2014). ATCOR consists of ATCOR 2 (used for flat terrain) and ATCOR 3 (used for rugged terrain, i.e., the 3 stands for 3 dimensions). ATCOR 4 is used with remotely sensed data acquired by suborbital remote sensing systems. The atmospheric correction algorithm employs the MODTRAN 4+ (Alder-Golden et al., 1999, 2005) radiative transfer code to calculate look-up tables (LUT) of the atmospheric correction functions (path radiance, atmospheric transmittance, direct and diffuse solar flux) that depend on scan angle, relative azimuth angle between scan line and solar azimuth, and terrain elevation. Examples of before and after atmospheric correction using ATCOR are shown in Figure 6-29. San and Suzen (2010), ERDAS ATCOR (2014) and Richter and Schlapfer (2014) provide detailed information about ATCOR.

Artifact Suppression Most robust radiative transfer-based atmospheric correction algorithms also suppress artifacts in the corrected remote sensor data. For example, FLAASH uses an adjustable spectral "polishing" based on Alder-Golden et al. (1999). This type of spectral polishing involves applying mild adjustments to reflectance data so that the spectra appear more like spectra of real materials as recorded on the ground by a handheld spectroradiometer. ENVI's EFFORT, ImSpec's ACORN, and Intergraph's ATCOR all have artifact suppression capability.

Single Spectrum Enhancement Radiative transfer-based atmospheric correction programs (e.g., ACORN, FLAASH) allow the user to incorporate *in situ* spectroradiometer measurements. This is commonly referred to as *single spectrum enhancement*. The analyst simply provides an accurate *in situ* spectral reflectance curve for a known homogeneous area (e.g., a bare soil field of kaolinite, a deep water body, or a parking lot) and also provides the geometric coordinates of this same location in the remotely sensed data. The atmospheric correction program then determines all of the traditional radiative transfer scattering and absorption adjustments, plus it attempts to make the pixel in the image have the same spectral characteristics as the *in situ* spectroradiometric data.

For example, an Analytical Spectral Devices (ASD) hand-held spectroradiometer (Figure 6-30a) was used to collect 46 *in situ* spectroradiometric measurements on the Savannah River Site Mixed Waste Management Facility (Figure 6-31a). GER DAIS 3715 hyperspectral data were collected and radiometrically corrected using ACORN and empirical line calibration (discussed in the next section). In addition, a single spectrum enhancement was applied to both methods. A comparison between the 46 *in situ* spectroradiometer measurements for each of the 35 bands with the four atmospheric correction treatments (ELC, ACORN, ELC+SSE, ACORN+ SSE) reveals that the ACORN and ELC with the single spectrum enhancement provided very accurate radiometric correction of the remote sensor data (Figure 6-31b). The DAIS 3715 sensor did not collect data in one of the optimum atmospheric absorption windows. This may be the reason why ACORN did not do as well as a stand-alone correction.

Absolute Atmospheric Correction Using Empirical Line Calibration

Absolute atmospheric correction may also be performed using **empirical line calibration** (ELC), which forces the remote sensing image data to match *in situ* spectral reflectance measurements, hopefully obtained at approximately the same time and on the same date as the remote sensing overflight. Empirical line calibration is based on the equation (Smith and Milton, 1999; ImSpec, 2014):

$$BV_k = \rho_\lambda A_k + B_k, \qquad (6.35)$$

where BV_k is the digital output value for a pixel in band k, ρ_λ equals the scaled surface reflectance of the materials within the remote sensor IFOV at a specific wavelength (λ), A_k is a multiplicative term affecting the BV, and B_k is an additive term. The multiplicative term is associated primarily with atmospheric transmittance and instrumental factors, and the additive term deals primarily with atmospheric path radiance and instrumental offset (i.e., dark current).

To use Empirical Line Calibration, the image analyst usually selects two or more areas in the scene with different albedos (e.g., one bright target such as sand and one dark target such as a deep, nonturbid water body).

ATCOR® Atmospheric Correction

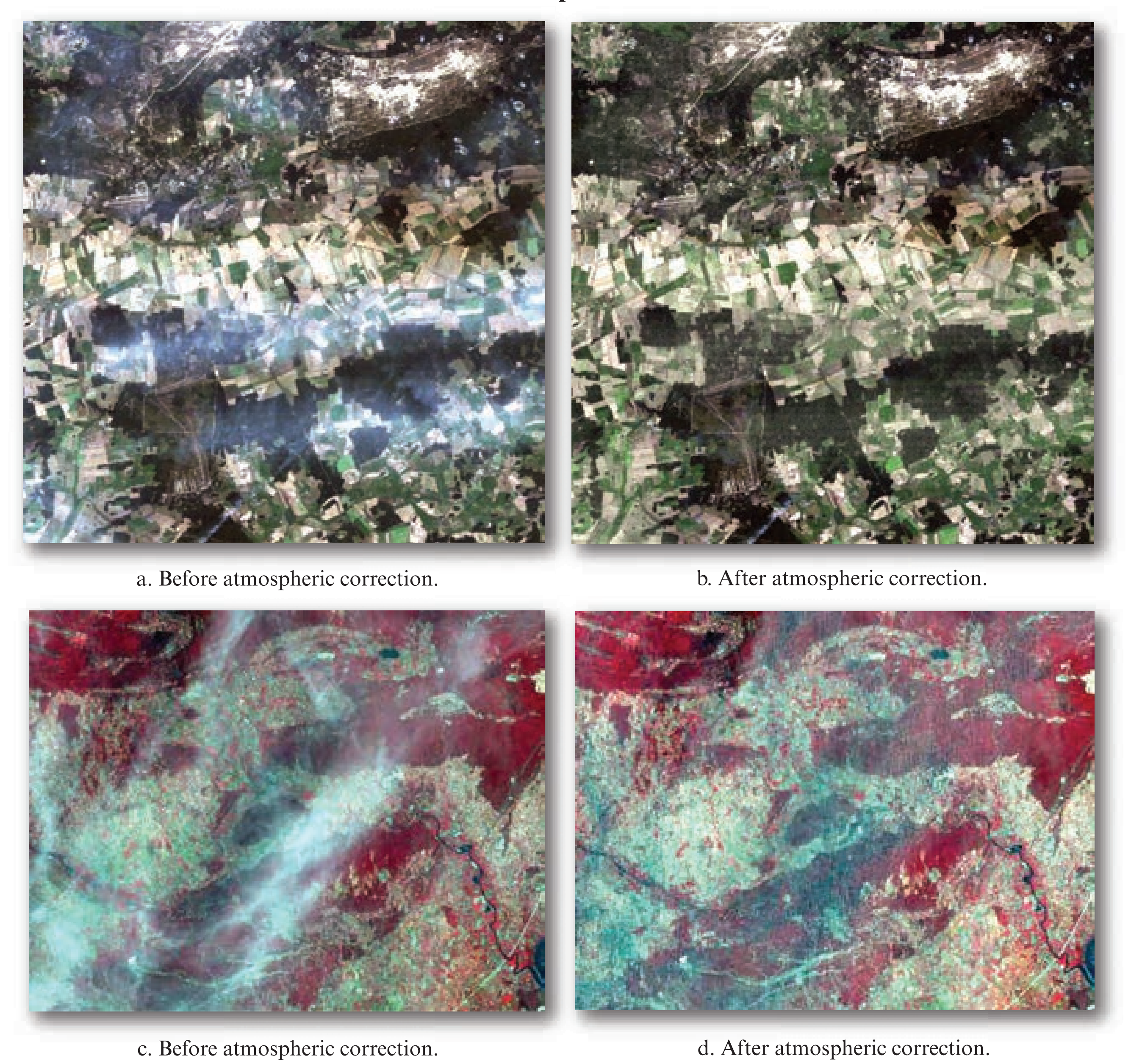


FIGURE 6–29 a) Natural color image containing substantial haze prior to atmospheric correction (Collwitz, Germany). b) Image after atmospheric correction using ATCOR. c) Color-infrared color composite containing substantial haze d) Image after atmospheric correction using ATCOR (images courtesy of Geosystems GmbH).

The areas selected should be as homogeneous as possible. *In situ* spectroradiometer measurements of these targets are made on the ground. The *in situ* and remote sensing–derived spectra are regressed and gain and offset values computed. The gain and offset values are then applied to the remote sensor data on a band by band basis, removing atmospheric attenuation. Note that the correction is applied band by band and not pixel by pixel, as it is with ACORN, FLAASH, and ATCOR.

To gain an appreciation of how empirical line calibration works, consider how GER DAIS 3715 radiance values of the Savannah River Site were converted to scaled surface reflectance values. Instead of trying to locate natural dark and light radiometric control points on the ground, dark and light 8 × 8 m calibration panel targets were located in the field at the time of a GER DAIS 3715 hyperspectral overflight (Figure 6-30b). Moran et al. (2001) provide guidance on how to prepare and maintain reference reflectance tarps. McCoy (2005) and Johannsen and Daughtry (2009)



a. *In situ* spectroradiometer calibration.



b. Calibration targets.

FIGURE 6–30 a) Field crew taking a spectroradiometer measurement from a calibrated reflectance standard on the tripod. b) 8 × 8 m black-and-white calibration targets at the Savannah River Site (Jensen et al., 2003).

provide additional information about how to collect *in situ* spectral reflectance data.

The spatial resolution of the imagery was 2.4 × 2.4 m. Image pixel brightness values for the targets were extracted from flightline 05 at two locations: dark (3,682,609 N; 438,864 E) and bright (3,682,608 N; 438,855 E). *In situ* ASD spectroradiometer measurements were obtained from the calibration targets. The remote sensing and *in situ* spectral measurements for the calibration targets were then paired and input to the empirical line calibration to derive the appropriate gain and offset values to atmospherically correct the hyperspectral data. Results of the empirical line calibration are summarized in Figure 6-31b.

It is important to note that most multispectral remote sensing datasets can be calibrated using empirical line calibration. The difficulty arises when trying to locate homogeneous bright and dark targets in the study, collecting representative *in situ* spectroradiometer measurements, and extracting uncontaminated pixels of the calibration targets from the imagery. If the analyst does not have access to *in situ* spectra obtained at the time of the remote sensing overflight, it might be possible to use spectra of such fundamental materials as clear water and sand (quartz) that are stored in spectral libraries.

Laboratory-derived spectra may be found at the ASTER Spectral Library (2014), which contains the: Johns Hopkins Spectral Library, the NASA Jet Propulsion Laboratory Spectral Library, and the U.S. Geological Survey spectral library (Baldridge et al., 2009). Hopefully, some of these materials exist in the scene and the analyst can locate the appropriate pixel and pair the image brightness values with the library *in situ* spectroradiometer data.

To demonstrate the use of empirical line calibration applied to historical imagery, consider the Landsat Thematic Mapper dataset shown in Figure 6-32a. *In situ* spectral reflectance measurements were not obtained when the Landsat TM imagery were acquired on February 3, 1994. Therefore, how can the brightness values in this historical imagery be converted into scaled surface reflectance? The answer lies in the use of empirical line calibration, *in situ* library spectra, and the Landsat TM brightness values ($BV_{i,j,k}$). This particular scene has a substantial amount of water and beach. Therefore, *in situ* library spectroradiometer data of water (from the Johns Hopkins library) and quartz (from the NASA JPL library) were utilized. The multispectral brightness values of one pixel of water and one pixel of beach were extracted from the Landsat TM data. These were paired with the *in situ* spectral reflectance measurements in the same six wavelengths (the Landsat 6 thermal infrared channel was not used). The empirical line calibration resulted in a Landsat TM dataset scaled to scaled surface reflectance. For example, consider a pixel of healthy loblolly pine. The original brightness value response is shown in Figure 6-32b. The scaled surface reflectance derived from the Landsat TM data is shown in Figure 6-32c. All pixels in the scene now represent scaled surface reflectance and not brightness value.

Scaled surface reflectance measurements obtained from this date of imagery can now be compared with scaled surface reflectance measurements extracted from other dates of Landsat TM imagery for purposes of monitoring biomass, etc. Of course, the use of library spectra is not as good as acquiring *in situ* spectroradiometry at selected sites in the scene at the time of the remote sensing data collection. Nevertheless, the technique is useful.

Relationship between 46 *In Situ* Spectroradiometer Measurements and Pixel Spectral Reflectance after Processing Using Empirical Line Calibration and ACORN

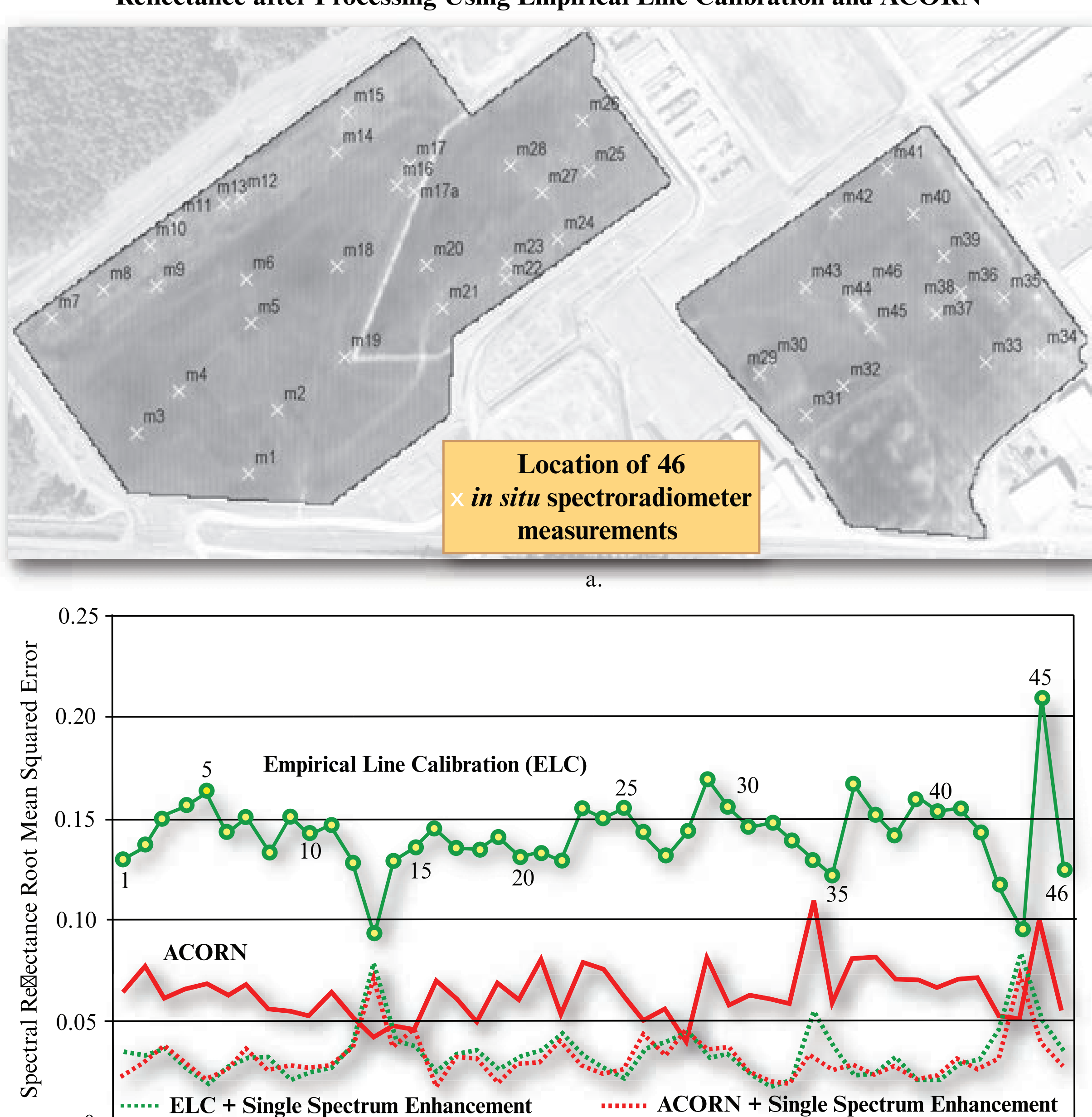


FIGURE 6–31 a) Forty-six *in situ* ASD spectroradiometer (400 – 2,500 nm) measurements were obtained on the Savannah River Site Mixed Waste Management Facility near Aiken, SC. b) The relationship between the 46 *in situ* spectroradiometer measurements and pixel spectral reflectance after processing using empirical line calibration (ELC) and ACORN. The ELC and ACORN with the single spectrum enhancements provided the most accurate atmospheric correction of the GER DAIS 3715 hyperspectral data (based on Jensen et al., 2003).

Relative Radiometric Correction of Atmospheric Attenuation

As previously discussed, absolute radiometric correction makes it possible to relate the digital counts in satellite or aircraft image data to scaled surface reflectance. Except when using empirical line calibration, this requires sensor calibration coefficients (to convert the original remote sensor data to W m^2 sr^{-1}) and an atmospheric correction algorithm usually based on radiative transfer code. Unfortunately, the application of these codes to a specific scene and date also requires knowledge of both the sensor spectral profile and the atmospheric properties at the time of remote sensor data collection (Du et al., 2002). If all of this type of information is available, then it is possible for

the atmospheric radiative transfer code to provide realistic estimates of the effects of atmospheric scattering and absorption and convert the imagery to scaled surface reflectance. If such information is not available, then relative radiometric correction techniques may be used.

Relative radiometric correction may be used 1) to *normalize* the intensities among the different bands within a single-date remotely sensed image, and 2) to *normalize* the intensities of bands of remote sensor data in multiple dates of imagery to a standard scene selected by the analyst.

Single-Image Normalization Using Histogram Adjustment

This simple method is based primarily on the fact that infrared data (>0.7 μm) are largely free of atmospheric scattering effects, whereas the visible region (0.4 – 0.7 μm) is strongly influenced by them. The method involves evaluating the histograms of the various bands of remotely sensed data of the desired scene. Normally, the data collected in the visible wavelengths (e.g., TM bands 1 to 3) have a higher minimum value because of the increased atmospheric scattering taking place in these wavelengths. For example, a histogram of Charleston, SC, TM data (Figure 6-33) reveals that it has a minimum of 51 and a maximum of 242. Conversely, atmospheric absorption subtracts brightness from the data recorded in the longer-wavelength intervals (e.g., TM bands 4, 5, and 7). This effect commonly causes data from the infrared bands to have minimums close to zero, even when few objects in the scene truly have a reflectance of zero (Figure 6-33).

If the histograms in Figure 6-33 are shifted to the left so that zero values appear in the data, the effects of atmospheric scattering will be somewhat minimized. This simple algorithm models the first-order effects of atmospheric scattering, or *haze.* It is based on a subtractive bias established for each spectral band. The bias may also be determined by evaluating a histogram of brightness values of a reference target such as deep water in all bands. The atmospheric effects correction algorithm is defined:

$$\text{output } BV_{i,j,k} = \text{input } BV_{i,j,k} - \text{bias} \qquad (6.36)$$

where input $BV_{i,j,k}$ = input pixel value at line i and column j of band k and output $BV_{i,j,k}$ = the adjusted pixel value at the same location.

In this example, the appropriate bias was determined for each histogram in Figure 6-33 and subtracted from the data. The histograms of the adjusted data are displayed in Figure 6-34. It was not necessary to adjust bands 5 and 7 for first-order atmospheric effects because they originally had minimums of zero. Hadjimitisis et al. (2010) used histogram adjustment (i.e., darkest pixel) atmospheric correction and found it of value for analyzing Landsat TM data for agricultural applications. Chrysoulakis et al. (2010) also found the simple histogram adjustment method to be useful.

Multiple-Date Image Normalization Using Regression

Multiple-date image normalization involves selecting a base image (b) and then transforming the spectral characteristics of all other images obtained on different dates (e.g., those obtained on dates b–1, b–2 and/or b+1, b+2, etc.) to have approximately the same radiometric scale as the base image. It is important to remember, however, that the radiometric scale used in a relative multiple-date image normalization will most likely be simple brightness values (e.g., BV with a range from 0 to 255) rather than scaled surface reflectance produced when conducting an absolute radiometric correction.

Multiple-date image normalization involves the selection of **pseudo-invariant features** (PIFs), often referred to as radiometric ground control points. To be of value in the multiple-date image normalization process, the pseudo-invariant features should have the following characteristics:

- The spectral characteristics of a PIF should change very little through time, although it is acknowledged that some change is inevitable. Deep nonturbid water bodies, bare soil, large rooftops, or other homogeneous features are candidates.
- The PIF should be at approximately the same elevation as the other land in the scene. Selecting a mountaintop PIF would be of little use in estimating atmospheric conditions near sea level because most aerosols in the atmosphere occur within the lowest 1,000 m.
- The PIF should normally contain only minimal amounts of vegetation. Vegetation spectral reflectance can change over time as a result of environmental stress and plant phenology. However, an extremely stable, homogeneous forest canopy imaged on approximate near-anniversary dates might be considered.
- The PIF must be in a relatively flat area so that incremental changes in Sun angle from date to date will have the same proportional increase or decrease in direct beam sunlight for all normalization targets.

Regression is used to relate the base image's PIF spectral characteristics with PIF spectra characteristics from other dates of imagery. The algorithm assumes



a. Emp irical line calibrated Landsat Thematic Mapper band image of Charleston, SC.

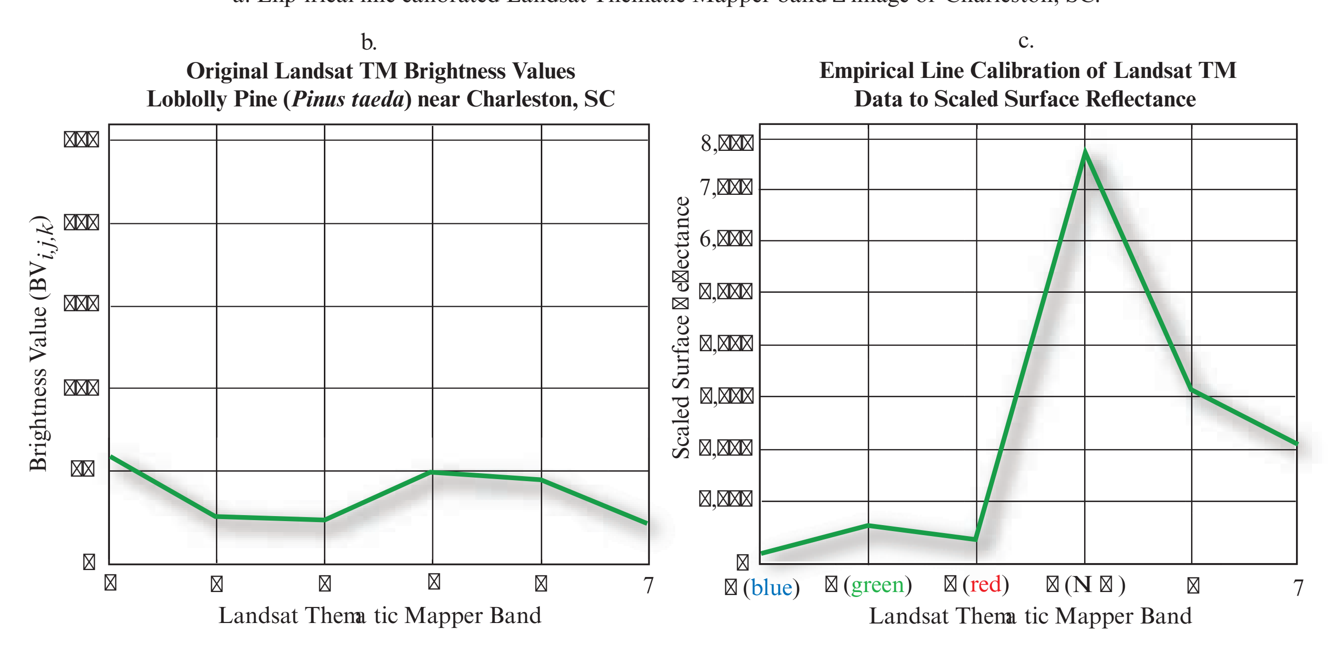


FIGURE 6–32 a) A Landsat Thematic Mapper image acquired on February 3, 1994, was radiometrically corrected using empirical line calibration and paired NASA JPL and Johns Hopkins University spectral library beach and water *in situ* spectroradiometer measurements and Landsat TM image brightness values ($BV_{i,j,k}$). Original imagery courtesy of NASA. b) A pixel of loblolly pine with its original brightness values in six bands (the TM band 6 thermal channel was not used). c) The same pixel after empirical line calibration to scaled surface reflectance. Note the correct chlorophyll absorption in the blue (band 1) and red (band 3) portions of the spectrum and the increase in near-infrared reflectance.

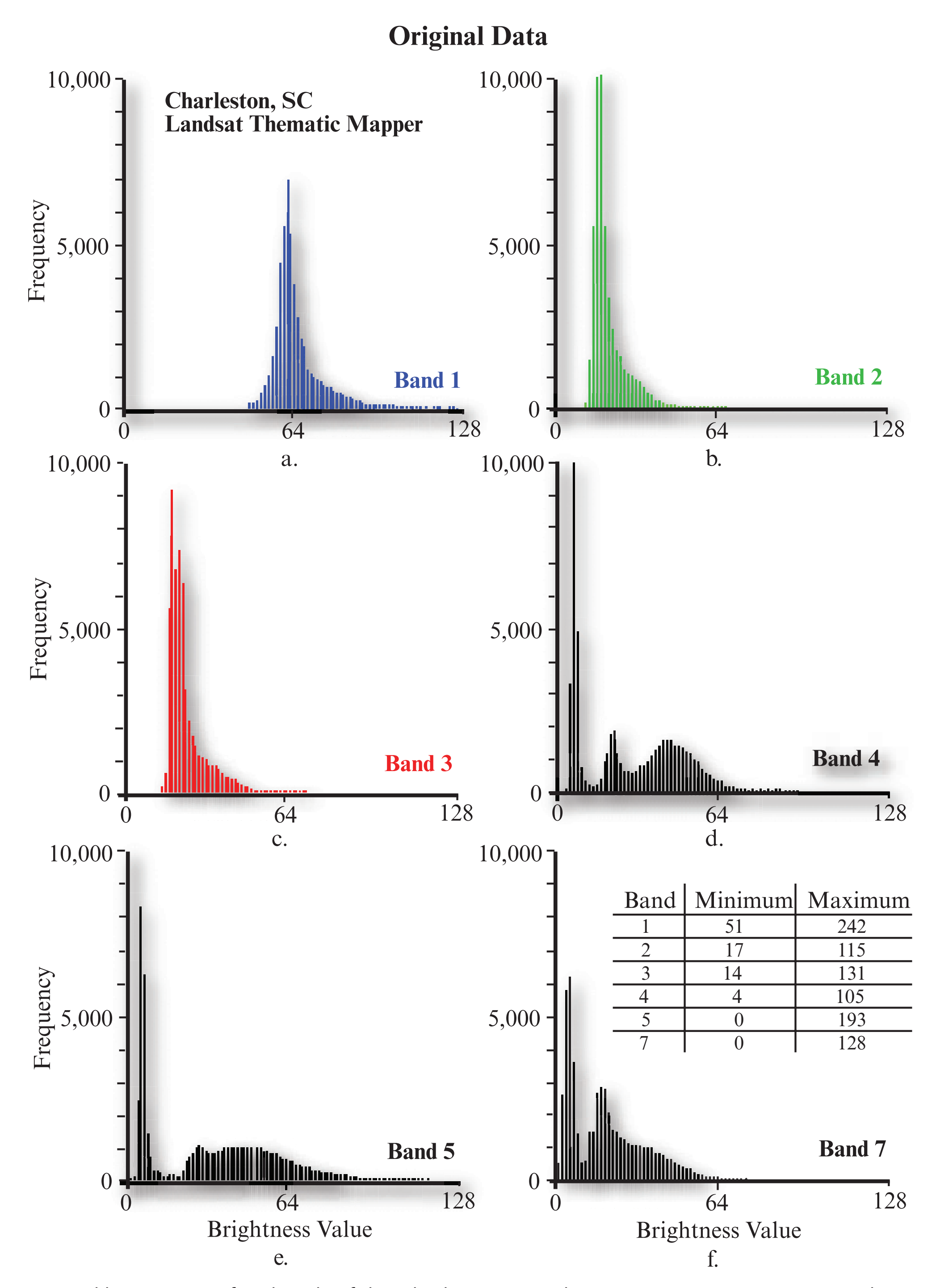


FIGURE 6–33 Original histograms of six bands of the Charleston, SC, Thematic Mapper scene. Atmospheric scattering in the visible regions has increased the minimum brightness values in bands 1, 2, and 3. Generally, the shorter the wavelengths sensed by each band, the greater the offset from a brightness value of zero.

that the pixels sampled at time $b + 1$ or $b - 1$ are linearly related to the pixels for the same locations on the base image (b). This implies that the spectral reflectance properties of the sampled pixels have not changed during the time interval. Thus, the key to the image regression method is the selection of quality pseudo-invariant features.

Numerous scientists have investigated the utility of pseudo-invariant features to normalize multiple-date imagery. For example, Caselles and Garcia (1989), Schott et al. (1988), and Hall et al. (1991) developed a *radiometric rectification technique* that corrected images of the same areas through the use of landscape elements whose reflectances were nearly constant over time. Jensen et al. (1995) used similar procedures to

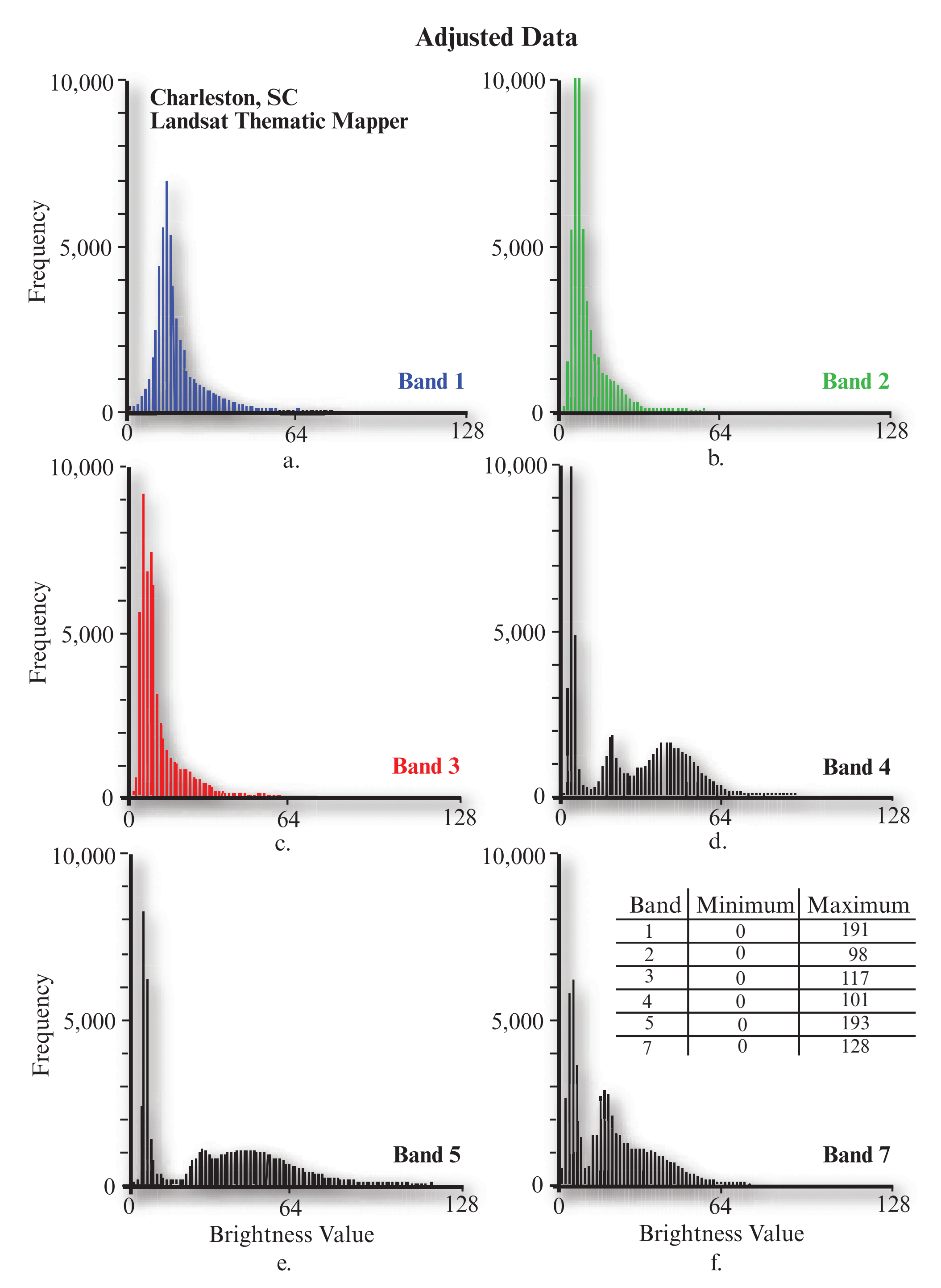


FIGURE 6–34 Results of applying a simple histogram adjustment atmospheric scattering correction to the data shown in Figure 6-33. Only the first four Thematic Mapper bands required the adjustment. This method does not correct for atmospheric absorption.

perform relative radiometric correction. Heo and Fitzhugh (2000) developed different techniques to select PIFs. Du et al. (2002) developed a procedure for the relative radiometric normalization of multitemporal satellite images where PIFs were selected objectively from images that were analyzed using principal component analysis (PCA).

A remote sensing study that identified changes in cattail distribution in the South Florida Water Management District can be used to demonstrate the radiometric normalization process (Jensen et al., 1995). Five predominantly cloud-free dates of satellite remote sensor data were collected by two sensor systems for Everglades Water Conservation Area 2A from 1973 to 1991. Landsat Multispectral Scanner (MSS) data were

TABLE 6–6 Characteristics of the remotely sensed satellite data used to inventory wetland in Water Conservation Area 2A of the South Florida Water Management District (Jensen et al., 1995).

Date	Type of Imagery	Bands Used	Nominal Instantaneous Field of View (m)	Rectification RMSE
3/22/73	Landsat MSS	1, 2, 4	79 × 79	–0.377
4/02/76	Landsat MSS	1, 2, 4	79 × 79	–0.275
10/17/82	Landsat MSS	1, 2, 4	79 × 79	–0.807
4/04/87	SPOT HRV	1, 2, 3	20 × 20	–0.675
8/10/91	SPOT HRV	1, 2, 3	20 × 20	–0.400

obtained in 1973, 1976, and 1982 and SPOT High Resolution Visible (HRV) multispectral (XS) data in 1987 and 1991. The specific date, type of imagery, bands used in the analysis, and nominal spatial resolution of the sensor systems are summarized in Table 6-6. Twenty (20) ground control points (GCPs) were obtained and used to rectify the August 10, 1991, remote sensor data to a standard map projection. The remotely sensed data were rectified to a Universal Transverse Mercator (UTM) map projection having 20 × 20 m pixels using a nearest-neighbor resampling algorithm and a root-mean-square error (RMSE) of – 0.4 pixel (Rutchey and Vilchek, 1994). All other images were resampled to 20 × 20 m pixels using nearest-neighbor resampling and registered to the 1991 SPOT data for change detection purposes. The RMSE statistic for each image is summarized in Table 6-6.

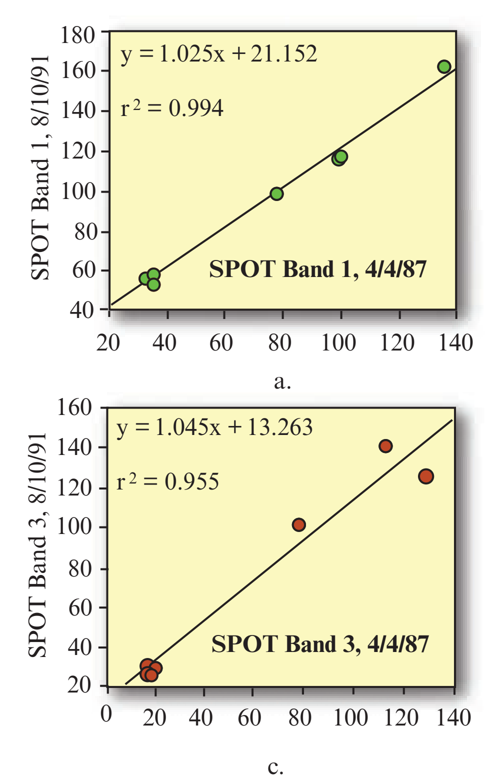
A problem associated with using historical remotely sensed data for change detection is that the data are usually from non-anniversary dates with varying Sun angle, atmospheric, and soil moisture conditions. The multiple dates of remotely sensed data should be normalized so that these effects can be minimized or eliminated (Hall et al., 1991).

Differences in direct-beam solar radiation due to variation in Sun angle and Earth-to-Sun distance can be calculated accurately, as can variation in pixel BVs due to detector calibration differences between sensor systems. However, removal of atmospheric and phase angle effects requires information about the gaseous and aerosol composition of the atmosphere and the bidirectional reflectance characteristics of elements within the scene (Eckhardt et al., 1990). Because atmospheric and bidirectional reflectance information were not available for any of the five scenes, an "empirical scene normalization" approach was used to match the detector calibration, astronomic, atmospheric, and phase angle conditions present in a reference scene. The August 10, 1991, SPOT HRV scene was selected as the base scene to which the 1973, 1976, 1982, and 1987 scenes were normalized. The 1991 SPOT image was selected because it was the only year for which quality *in situ* ground reference data were available.

Image normalization was achieved by applying regression equations to the 1973, 1976, 1982, and 1987 imagery to predict what a given *BV* would be if it had been acquired under the same conditions as the 1991 reference scene. These regression equations were developed by correlating the brightness of pseudo-invariant features present in both the scene being normalized and the reference (1991) scene. PIF were assumed to be constant reflectors, so any changes in their brightness values were attributed to detector calibration, and astronomic, atmospheric, and phase angle differences. Once these variations were removed, changes in *BV* could be related to changes in surface conditions.

Multiple wet (water) and dry (e.g., unvegetated bare soil) PIF were found in the base year image (1991) and each of the other dates of imagery (e.g., 1987 SPOT data). A total of 21 radiometric control points were used to normalize the 1973, 1976, 1982, and 1987 data to the 1991 SPOT data. It is useful to summarize the nature of the normalization targets used and identify adjustments that had to be made when trying to identify dry soil targets in a humid subtropical environment. Radiometric normalization targets found within the 1987 and 1991 SPOT data consisted of three wet points obtained just to the north of WCA-2A within WCA-1 and three dry points extracted from an excavated area, a dry lake area, and a limestone road area. The brightness values of the early image targets (e.g., 1987) were regressed against the brightness values of the base image targets (e.g., 1991) for each band (Figure 6-35a–c). The coefficients and intercept of the equation were used to compute a normalized 1987 SPOT dataset, which had approximately the same spectral characteristics as the 1991 SPOT data. Each regression model contained an additive component that corrected for

Radiometric Normalization of Multiple-date South Florida SPOT Data



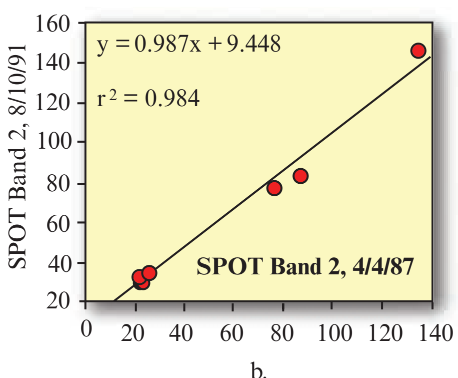


FIGURE 6–35 a) Relationship between the same wet and dry regions found in both the 4/4/87 and 8/10/91 SPOT band 1 (green) dataset. The equation was used to normalize the 4/4/87 data to the 8/10/91 SPOT data as per methods described in Eckhardt et al. (1990). b) Relationship between wet and dry regions found in both the 4/4/87 and 8/10/91 SPOT band 2 (red) dataset. c) Relationship between wet and dry regions found in both the 4/4/87 and 8/10/91 SPOT band 3 (near-infrared) dataset (Jensen, J. R., Rutchey, K., Koch, M. and S. Narumalani,1995, "Inland Wetland Change Detection in the Everglades Water Conservation Area 2A Using a Time Series of Normalized Remotely Sensed Data," *Photogrammetric Engineering & Remote Sensing*, 61(2):199–209. Reprinted with permission of the American Society for Photogrammetry & Remote Sensing).

the difference in atmospheric path radiance among dates and a multiplicative term that corrected for the difference in detector calibration, Sun angle, Earth–Sun distance, atmospheric attenuation, and phase angle between dates.

The 1982 MSS data were normalized to the 1991 data using (1) three common wet targets found within WCA-1 and (2) two dry points extracted from a bare soil excavation area in 1982, which progressed northward about 300 m (15 pixels) in the y dimension by 1991 (i.e., the x dimension was held constant). Thus, two noncommon dry radiometric control points were extracted for this date. Hall et al. (1991) suggest that the members of the radiometric control sets may not be the same pixels from image to image, in contrast to geometric control points for spatial image rectification, which are composed of identical elements in each scene. Furthermore, they suggest that "using fixed elements inevitably requires manual selection of sufficient numbers of image-to-image pairs of suitable pixels, which can be prohibitively labor intensive, particularly when several images from a number of years are being considered." Such conditions were a factor in the Everglades study.

The 1976 MSS data were normalized to the 1991 data using three wet targets located in WCA-1 and two dry points extracted along a bare soil road and a limestone bare soil area. The 1973 MSS data were normalized to the 1991 data using two wet and three dry targets. The greater the time between the base image (e.g., 1991) and the earlier year's image (e.g., 1973), the more difficult it is to locate unvegetated, dry normalization targets. For this reason, analysts sometimes use synthetic, pseudo-invariant features such as concrete, asphalt, rooftops, parking lots, and roads when normalizing historic remotely sensed data (Schott et al., 1988; Caselles and Garcia, 1989; Hall et al., 1991).

The normalization equations for each date are summarized in Table 6-7. The gain (slope) associated with the SPOT data was minimal, while the historical MSS data required significant gain and bias adjustments (because some MSS data were not originally acquired as 8-bit data). The methodology applied to all images minimized the differences in Sun angle, atmospheric effects, and soil moisture conditions between the dates. The radiometrically corrected remote sensor data were then classified and used to monitor wetland change (Jensen et al., 1995).

The ability to use remotely sensed data to classify land cover accurately is contingent on there being a robust relationship between remotely sensing brightness value (BV) and actual surface conditions. However, factors such as Sun angle, Earth–Sun distance, detector calibration differences among the various sensor systems,

TABLE 6–7 Equations used to normalize the radiometric characteristics of the historic remote sensor data with the August 10, 1991, SPOT XS data (Jensen et al., 1995).

Date	Band	Slope	*y*-intercept	r^2
3/22/73	MSS 1	1.40	31.19	0.99[a]
	2	1.01	23.49	0.98
	4	3.28	23.48	0.99
4/02/76	MSS 1	0.57	31.69	0.99
	2	0.43	21.91	0.98
	4	3.84	26.32	0.96
10/17/82	MSS 1	2.52	16.117	0.99
	2	2.142	8.488	0.99
	4	1.779	17.936	0.99
4/04/87	SPOT 1	1.025	21.152	0.99
	2	0.987	9.448	0.98
	3	1.045	13.263	0.95

[a]All regression equations were significant at the 0.001 level.

atmospheric condition, and Sun–target–sensor (phase angle) geometry will affect pixel brightness value. Image normalization reduces pixel *BV* variation caused by nonsurface factors, so that variations in pixel brightness value among dates may be related to actual changes in surface conditions. Normalization may allow the use of pixel classification logic developed from a base year scene to be applied to the other normalized scenes.



Correcting for Slope and Aspect Effects

The previous sections discussed how scattering and absorption in the atmosphere can attenuate the radiant flux recorded by the sensor system. Topographic slope and aspect may also introduce radiometric distortion of the recorded signal. In some locations, the area of interest might even be in complete shadow, dramatically affecting the brightness values of the pixels involved. For these reasons, research has been directed toward the removal of topographic effects, especially in mountainous regions, on Landsat and SPOT digital multispectral data (Shepherd et al., 2014). The goal of a slope-aspect correction is to remove topographically induced illumination variation so that two objects having the same reflectance properties show the same brightness value in the image despite their different orientation to the Sun's position. If the topographic slope-aspect correction is applied effectively, the three-dimensional impression we get when looking at a satellite image of mountainous terrain should be somewhat subdued. A good slope-aspect correction is believed to improve forest stand classification when compared to noncorrected imagery (Civco, 1989; Meyer et al., 1993).

Teillet et al. (1982) described four topographic slope-aspect correction methods: the simple cosine correction, two semi-empirical methods (the Minnaert method and the *C* correction), and a statistic-empirical correction. Each correction is based on *illumination,* which is defined as the cosine of the incident solar angle, thus representing the proportion of the direct solar radiation hitting a pixel. The amount of illumination is dependent on the relative orientation of the pixel toward the Sun's actual position (Figure 6-36). Each slope-aspect topographic correction method to be discussed requires a digital elevation model (DEM) of the study area. The DEM and satellite remote sensor data (e.g., Landsat TM data) must be geometrically registered and resampled to the same spatial resolution (e.g., 30 × 30 m pixels). The DEM is processed so that each pixel's brightness value represents the amount of illumination it should receive from the Sun. This information is then modeled using one of the four algorithms to enhance or subdue the original brightness values of the remote sensor data.

The Cosine Correction

The amount of irradiance reaching a pixel on a slope is directly proportional to the cosine of the incidence angle *i*, which is defined as the angle between the normal on the pixel in question and the zenith direction (Teillet et al., 1982). This assumes 1) Lambertian surfaces, 2) a constant distance between Earth and the Sun, and 3) a constant amount of solar energy illuminating Earth (somewhat unrealistic assumptions!). Only the part $\cos i$ of the total incoming irradiance, E_g, reaches the inclined pixel. It is possible to perform a simple topographic slope-aspect correction of the remote sensor data using the following cosine equation:

$$L_H = L_T \frac{\cos\theta_o}{\cos i} \tag{6.37}$$

where

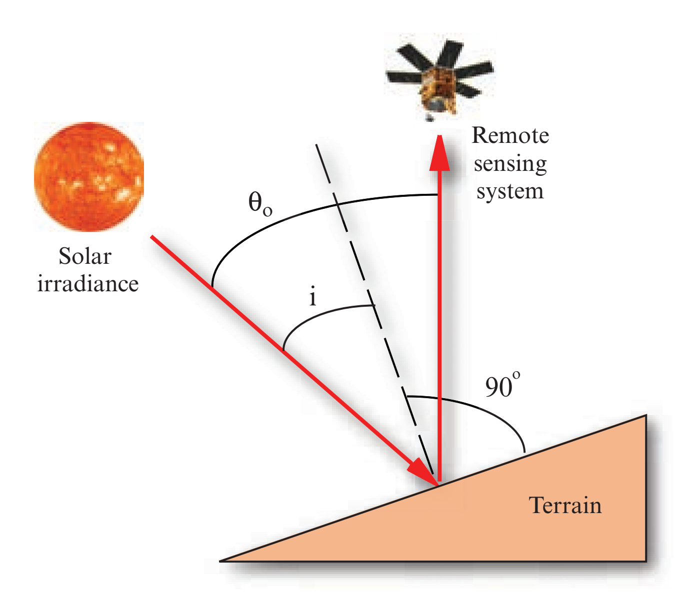


FIGURE 6–36 Representation of the Sun's angle of incidence, *i*, and the solar zenith angle, θ_o.

L_H = radiance observed for a horizontal surface (i.e., slope-aspect-corrected remote sensor data)

L_T = radiance observed over sloped terrain (i.e., the raw remote sensor data)

θ_o = Sun's zenith angle

i = Sun's incidence angle in relation to the normal on a pixel (Figure 6-36).

Unfortunately, this method models only the direct part of the irradiance that illuminates a pixel on the ground. It does not take into account diffuse skylight or light reflected from surrounding mountainsides that may illuminate the pixel in question. Consequently, weakly illuminated areas in the terrain receive a disproportionate brightening effect when the cosine correction is applied. Basically, the smaller the cos *i*, the greater the overcorrection is (Meyer et al., 1993). Nevertheless, several researchers have found the cosine correction of value. For example, Civco (1989) achieved good results with the technique when used in conjunction with empirically derived correction coefficients for each band of imagery.

The Minnaert Correction

Teillet et al. (1982) introduced the Minnaert correction to the basic cosine function:

$$L_H = L_T\left(\frac{\cos\ \theta_o}{\cos\ i}\right)^k \qquad (6.38)$$

where k = the Minnaert constant.

The constant varies between 0 and 1 and is a measure of the extent to which a surface is Lambertian. A perfectly Lambertian surface has $k = 1$ and represents a traditional cosine correction. Meyer et al. (1993) describe how k may be computed empirically. Lu et al. (2008) developed a pixel-based Minnaert correction method which reduced topographic effects when applied to Landsat ETM$^+$ data over mountainous terrain.

A Statistical–Empirical Correction

For each pixel in the scene, it is possible to correlate (1) the predicted illumination ($\cos i \times 100$) from the DEM with (2) the actual remote sensor data. For example, Meyer et al. (1993) correlated Landsat TM data of known forest stands in Switzerland with the predicted illumination from a high-resolution DEM. Any slope in the regression line suggests that a constant type of forest stand will appear differently on different terrain slopes. Conversely, by taking into account the statistical relationship in the distribution, the regression line may be rotated based on the following equation:

$$L_H = L_T - \cos(i)m - b + \overline{L_T} \qquad (6.39)$$

where

L_H = radiance observed for a horizontal surface (i.e., slope-aspect-corrected remote sensor data)

L_T = radiance observed over sloped terrain (i.e., the raw remote sensor data)

$\overline{L_T}$ = average of L_T for forested pixels (according to ground reference data)

i = Sun's incidence angle in relation to the normal on a pixel (Figure 6-36)

m = slope of the regression line

b = y-intercept of the regression line.

Application of this equation makes a specific object (e.g., a particular type of deciduous forest) independent of cos *i* and produces the same brightness values (or radiance) throughout the image for this object.

The C Correction

Teillet et al. (1982) introduced an additional adjustment to the cosine function called the *c* correction:

$$L_H = L_T\frac{\cos\ \theta_o + c}{\cos\ i + c} \qquad (6.40)$$

where

$c = \frac{b}{m}$ in the previous regression equation.

Similar to the Minnaert constant, c increases the denominator and weakens the overcorrection of faintly illuminated pixels.

Local Correlation Filter

Shepherd et al. (2014) used a local correlation filter applied to a panchromatic band of imagery and a digital elevation model of the same area. The 7×7 spatial moving filter determined the offset between a digital elevation model (DEM) and ortho-rectified satellite imagery for every pixel. The mean correlation coefficient between the input panchromatic band and the DEM within the moving 7×7 pixel window was increased to 0.71 by using the local correlation filter. They suggested their correction removed the majority of the high-frequency artifacts.



References

Agrawal, G., and J. Sarup, 2011, "Comparison of QUAC and FLAASH Atmospheric Correction Modules on EO-1 Hyperion Data of Sanchi," *International Journal of Advanced Engineering Sciences and Technologies*, 4(1):178-186.

Alder-Golden, S. M., Acharya, P. K., Berk, A., Matthew, M. W., and D. Gorodetzky, 2005, "Remote Bathymetry of the Littoral Zone from AVIRIS, LASH, and QuickBird Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 43(2): 337–347.

Alder-Golden, S. M., Matthew, M. W., Bernstein, L. S., Levine, R. Y., Berk, A., Richtsmeier, S. C., Acharya, P. K., Anderson, G. P., Felde, G., Gardner, J., Hoke, M., Jeong, L. S., Pukall, B., Mello, J., Ratkowski, A., and H. H. Burke, 1999, "Atmospheric Correction for Short-wave Spectral Imagery Based on MODTRAN4," *SPIE Proceedings on Imaging Spectrometry V*, 3753:61–69.

ASTER Spectral Library, 2014, contains the *Johns Hopkins University Spectral Library, Jet Propulsion Laboratory Spectral Library,* and the *U.S. Geological Survey Spectral Library*, Pasadena: NASA JPL, http://speclib.jpl.nasa.gov/.

Baldridge, A. M., Hook, S. J., Grove, C.I., and G. Rivera, 2009, "The ASTER Spectral Library Version 2.0," *Remote Sensing of Environment*, 113:711–715.

Bernstein, L. S., Alder-Golden, S. M., Sundberg, R., and A. Ratkowski, 2008, "In-scene-based Atmospheric Correction of Uncalibrated VISible-SWIR (VIS-SWIR) Hyper- and Multispectral Imagery," *Proceedings*, Europe Security and Defense, Remote Sensing, Volume 7107 (2008), 8 p.

Caselles, V., and M. J. Garcia, 1989, "An Alternative Simple Approach to Estimate Atmospheric Correction in Multitemporal Studies," *International Journal of Remote Sensing*, 10(6):1127–1134.

Chrysoulakis, N., Abrams, M., Feidas, H., and K. Arai, 2010, "Comparison of Atmospheric Correction Methods using ASTER Data for the Area of Crete: the ATMOSAT Project," *International Journal of Remote Sensing*, 31:6347–6385.

Civco, D. L., 1989, "Topographic Normalization of Landsat Thematic Mapper Digital Imagery," *Photogrammetric Engineering & Remote Sensing*, 55(9):1303–1309.

Colwell, R. N., (Ed.), 1983, *Manual of Remote Sensing*, 2nd ed., Falls Church, VA: American Society of Photogrammetry, 2440 p.

Congalton, R. G., and K. Green, 2009, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Boca Raton, FL: Lewis Publishers, 183 p.

Cracknell, A. P., and L. W. Hayes, 1993, "Atmospheric Corrections to Passive Satellite Remote Sensing Data," Chapter 8 in *Introduction to Remote Sensing*, London: Taylor & Francis, 116–158.

Du, Y., Teillet, P., and J. Cihlar, 2002, "Radiometric Normalization of Multitemporal High-resolution Satellite Images with Quality Control for Land Cover Change Detection," *Remote Sensing of Environment*, 82:123.

Eckhardt, D. W., Verdin, J. P., and G. R. Lyford, 1990, "Automated Update of an Irrigated Lands GIS Using SPOT HRV Imagery," *Photogrammetric Engineering & Remote Sensing*, 56(11):1515–1522.

ERDAS ATCOR, 2014, *ATCOR for ERDAS Imaging — Atmospheric Correction for Professionals*, http://www.geosytems.de/atcor/index.html.

Exelis Atmospheric Correction, 2014, *About the ENVI Atmospheric Correction Module*, Boulder: Exelis, Inc., http://www.exelisvis.com/docs/AboutAtmosphericCorrectionModule.html.

Exelis FLAASH Hyperspectral, 2014, *ENVI Classic Tutorial: Atmospherically Correcting Hyperspectral Data Using FLAASH*, Boulder: Exelis, Inc., 8 p., http://www.exelisvis.com/portals/0/pdfs/envi/FLAASH_Hyperspectral.pdf.

Exelis FLAASH Multispectral, 2014, *ENVI Classic Tutorial: Atmospherically Correcting Multispectral Data Using FLAASH*, Boulder: Exelis, Inc., 14 p., http://www.exelisvis.com/portals/0/pdfs/envi/FLAASH_Multispectral.pdf.

Exelis QUAC, 2014, *QUick Atmospheric Correction (QUAC)*, Boulder: Exelis, Inc., http://www.exelisvis.com/docs/QUAC.html and http://www.exelisvis.com/docs/BackgroundQUAC.html

Felde, G. W., Anderson, G. P., Alder-Golden, S. M., Matthew, M., and A. Berk, 2003, "Analysis of Hyperion Data with the FLAASH Atmospheric Correction Algorithm: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery," *SPIE Aerosense Conference*, Orlando, 21–25, April 2003.

Gao, B. C., Montes, M. J., Davis, C. O., and A. F. H. Goetz, 2009, "Atmospheric Correction Algorithms for Hyperspectral Remote Sensing of Land and Ocean," *Remote Sensing of Environment*, 113:S17–S24.

German Aerospace Center, 2014, *Atmospheric Correction,* German Aerospace Center, http://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-5450/10028_read-20715/.

Green, R. O., 2003, "Introduction to Atmospheric Correction," Chapter 2 in *ACORN Tutorial*, Boulder: Analytical Imaging and Geophysics, LLC, 12–18.

Hadjimitisis, D. G., Papadavid, G., Agapiou, A., Themistocleous, K., Hadjimitisis, M. G., Retalis, A., Michaelides, S., Chrysoulakis, N., Toulios, L., and C. R. Clayton, 2010, "Atmospheric Correction for Satellite Remotely Sensed Data Intended for Agricultural Applications: Impact on Vegetation Indices," *Natural Hazards Earth System Science*, 10:89–95.

Hall, F. G., Strebel, D. E., Nickeson, J. E., and S. J. Goetz, 1991, "Radiometric Rectification: Toward a Common Radiometric Response Among Multidate, Multisensor Images," *Remote Sensing of Environment*, 35:11–27.

He, Q., and C. Chen, 2014, "A New Approach for Atmospheric Correction of MODIS Imagery in Turbid Coastal Waters: A Case Study for the Pearl River Estuary, *Remote Sensing Letters,* 5(3):249–257.

Heo, J. and F. W. Fitzhugh, 2000, "A Standardized Radiometric Normalization Method for Change Detection Using Remotely Sensed Imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, 60:173–181.

ImSpec ACORN, 2014, *ACORN 6.1x,* Boulder: ImSpec, LLC, http://www.imspec.com/.

Jensen, J. R., and R. R. Jensen, 2013, *Introductory Geographic Information Systems*, Boston: Pearson, 400 p.

Jensen, J. R., Botchway, K., Brennan-Galvin, E., Johannsen, C., Juma, C., Mabogunje, A., Miller, R., Price, K., Reining, P., Skole, D., Stancioff, A., and D. R. Taylor, 2002, *Down to Earth: Geographic Information for Sustainable Development in Africa*, Washington: National Academy Press, 155 p.

Jensen, J. R., Cowen, D., Narumalani, S., Weatherbee, O., and J. Althausen, 1993, "Evaluation of CoastWatch Change Detection Protocol in South Carolina," *Photogrammetric Engineering & Remote Sensing*, 59(6):1039–1046.

Jensen, J. R., Hadley, B. C., Tullis, J. A., Gladden, J., Nelson, S., Riley, S., Filippi, T., and M. Pendergast, 2003, *2002 Hyperspectral Analysis of Hazardous Waste Sites on the Savannah River Site*, Aiken, SC: Westinghouse Savannah River Company, WSRC-TR-2003-0025, 52 p.

Jensen, J. R., Rutchey, K., Koch, M., and S. Narumalani, 1995, "Inland Wetland Change Detection in the Everglades Water Conservation Area 2A Using a Time Series of Normalized Remotely Sensed Data," *Photogrammetric Engineering & Remote Sensing*, 61(2):199–209.

Johannsen, C. J., and C. S. T. Daughtry, 2009, "Chapter 17: Surface Reference Data Collection," in Warner, T. A., Nellis, M. D. and G. M. Foody (Eds.), *The Sage Handbook of Remote Sensing*, Los Angeles: Sage Publications, 244–256.

Kaufman, Y. J., Wald, A. E., Remer, L. A., Gao, B. C., Li, R. R., and F. Flynn, 1997, "The MODIS 2.1-mm Channel Correlation with Visible Reflectance for Use in Remote Sensing of Aerosol," *IEEE Transactions on Geoscience and Remote Sensing*, 35:1286–1298.

Konecny, G., 2014, *Geoinformation: Remote Sensing, Photogrammetry and Geographic Information Systems*, Boca Raton: CRC Press, 416 p.

Kruse, F. A., 2004, "Comparison of ATREM, ACORN, and FLAASH Atmospheric Corrections using Low-Altitude AVIRIS Data of Boulder, Colorado," *Proceedings 13th JPL Airborne Geoscience Workshop*, Jet Propulsion Laboratory, 31 March – 2 April 2004, Pasadena, CA, JPL Publication 05-3, at ftp://popo.jpl.nasa.gov/pub/docs/workshops/04_docs/Kruse-JPL2004_ATM_Compare.pdf.

Lu, D., Ge, H., He, S., Xu, A., Zhou, G., and H. Du, 2008, "Pixel-based Minnaert Correction for Reducing Topographic Effects on a Landsat ETM^+ Image," *Photogrammetric Engineering & Remote Sensing*, 74(11):1343–1350.

Matthew, M., Alder-Golden, S., Berk, A., Felde, G., Anderson, G., Gorodestzky, D., Paswaters, S., and M. Shippert, 2003, "Atmospheric Correction of Spectral Imagery: Evaluation of the FLAASH Algorithm with AVIRIS Data," S*PIE Proceeding, Algorithm and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*.

Matthew, M., Alder-Golden, S., Berk, A., Richtsmeier, S., Levin, R., Bernstein, L., Acharya, P., Anderson, G., Felde, G., Hoke, M., Ratkowski, A., Burke, H., Kaiser, R., and D. Miller, 2000, "Status of Atmospheric Correction Using a MODTRAN4-based Algorithm," *SPIE Proceedings Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, 4049:199–207.

McCoy, R., 2005, *Field Methods in Remote Sensing*, NY: Guilford, 159 p.

Meyer, P., Itten, K. I., Kellenberger, T., Sandmeier, S., and R. Sandmeier, 1993, "Radiometric Corrections of Topographically Induced Effects on Landsat TM Data in an Alpine Environment," *ISPRS Journal of Photogrammetry and Remote Sensing* 48(4):17–28.

Miller, S. W., and E. Vermote, 2002, *NPOESS Visible/Infrared Imager/Radiometer Suite: Algorithm Theoretical Basis Document*, Version 5, Lanham, MD: Raytheon, 83 p.

Milman, A. S., 1999, *Mathematical Principles of Remote Sensing: Making Inferences from Noisy Data*, Ann Arbor, MI: Ann Arbor Press, 37 p.

Moran, S. M., Bryant, R. B., Clarke, T. R., and J. Qi, 2001, "Deployment and Calibration of Reference Reflectance Tarps for Use with Airborne Imaging Sensors," *Photogrammetric Engineering & Remote Sensing*, 67(3):273–286.

Nassau, K., 1983, *The Physics and Chemistry of Color: The Fifteen Causes of Color*, NY: John Wiley & Sons.

Nassau, K., 1984, "The Physics of Color," in *Science Year 1984*, Chicago: World Book, 126–139.

Pedrotti, L. S., 2014, "Basic Geometrical Optics," in *Fundamentals of Photonics*, SPIE International Society for Optics and Photonics, 1–44, available free from http://spie.org/x17229.xml.

Ready, J., 2014, "Optical Detectors and Human Vision," in *Fundamentals of Photonics*, SPIE International Society for Optics and Photonics, 1–38, available free from http://spie.org/x17229.xml.

Richards, J. A., 2013, *Remote Sensing Digital Image Analysis*, 5th Ed., New York: Springer-Verlag, 494 p.

Richter, R., and D. Schlapfer, 2014, *Atmospheric / Topographic Correction for Airborne Imagery: ATCOR-4 User Guide*, DLR – German Aerospace Center, 226 p., http://www.rese.ch/pdf/atcor4_manual.pdf.

Rutchey, K., and L. Vilchek, 1994, "Development of an Everglades Vegetation Map Using a SPOT Image and the Global Positioning System," *Photogrammetric Engineering & Remote Sensing*, 60(6):767–775.

San, B. T., and M. L. Suzen, 2010, "Evaluation of Different Atmospheric Correction Algorithms for EO-1 Hyperion Imagery," *Intl. Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, 38(8):392–397.

Schott, J. R., Salvaggio, C., and W. J. Wolchok, 1988, "Radiometric Scene Normalization Using Pseudoinvariant Features," *Remote Sensing of Environment*, 26:1–16.

Seeber, F., 2014, "Light Sources and Laser Safety," in *Fundamentals of Photonics*, SPIE International Society for Optics and Photonics, 1–34, available free from http://spie.org/x17229.xml.

Shepherd, J. D., Dymond, J. R., Gillingham, S., and P. Bunting, 2014, "Accurate Registration of Optical Satellite Imagery with Elevation Models for Topographic Correction," *Remote Sensing Letters*, 5(7):637–641.

Slater, P. N., 1980, *Remote Sensing Optics and Optical Systems*. Reading, MA: Addison-Wesley, 575 p.

Smith, G. M., and E. J. Milton, 1999, "The Use of Empirical Line Method to Calibrate Remotely Sensed Data to Reflectance," *International Journal of Remote Sensing*, 20:2653–2662.

Song, C., Woodcock, C. E., Soto, K. C., Lenney, M. P. and S. A. Macomber, 2001, "Classification and Change Detection Using Landsat TM Data: When and How to Correct Atmospheric Effects?" *Remote Sensing of Environment*, 75:230–244.

Strahler, A. N., and A. H. Strahler, 1989, *Elements of Physical Geography*, 4th ed., New York: John Wiley & Sons, 562 p.

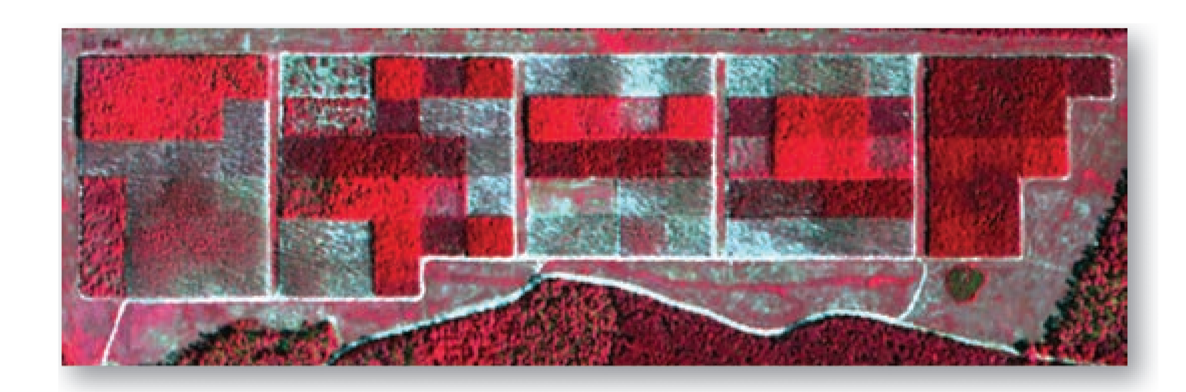
Teillet, P. M., Guindon, B., and D. G. Goodenough, 1982, "On the Slope-aspect Correction of Multispectral Scanner Data," *Canadian Journal of Remote Sensing*, 8(2):84–106.

Vandergriff, L. J., 2014, "Module 1.1: Nature and Properties of Light," in *Fundamentals of Photonics*, SPIE International Society for Optics and Photonics, 1–38, available free from http://spie.org/x17229.xml.

Vermote, E., Tanre, D., Deuze, J. L., Herman, M., and J. J. Morcrette, 1997, *Second Simulation of the Satellite Signal in the Solar Spectrum (6S)*, Code 923, Washington: NASA Goddard Space Flight Center, 54 p.

Vermote, E. F., El Saleous, N. Z., and C. O. Justice, 2002, "Atmospheric Correction of MODIS Data in the Visible to Middle Infrared: First Results," *Remote Sensing of Environment*, 83:97–111.

7 Geometric Correction



It would be wonderful if every remotely sensed image contained data that were already in their proper geometric *x, y* locations. This would allow each image to be used as if it were a map. Unfortunately, this is not the case. Instead, it is usually necessary to **preprocess** the remotely sensed data and remove the geometric distortion so that individual picture elements are in their proper planimetric (*x, y*) map locations (Purkis and Klemas, 2011). This allows remote sensing–derived information to be related to other thematic data in geographic information systems (GIS) (Merchant and Narumalani, 2009; Jensen and Jensen, 2013) or spatial decision support systems (SDSS) (Marcomini et al., 2008; Sugumaran and DeGrotte, 2010). Geometrically corrected imagery can be used to extract accurate distance, polygon area, and direction (bearing) information (Gonzalez and Woods, 2007; Wolf et al., 2013). Geometrically inaccurate remote sensor data can have very serious impacts on remote sensing–derived products (McRoberts, 2010).



Internal and External Geometric Error

Remotely sensed imagery typically exhibits *internal* and *external geometric error*. It is important to recognize the source of the internal and external error and whether it is *systematic* (predictable) or *nonsystematic* (random). Systematic geometric error is generally easier to identify and correct than random geometric error.

Internal Geometric Error

Internal geometric errors are generally introduced by the remote sensing system itself or in combination with Earth rotation or curvature characteristics. These distortions are often systematic (predictable) and may be identified and then corrected using prelaunch or in-flight platform ephemeris (i.e., information about the geometric characteristics of the sensor system and the Earth at the time of data acquisition). Geometric distortions in imagery that can sometimes be corrected through analysis of sensor characteristics and ephemeris data include:

- skew caused by Earth rotation effects,
- scanning system–induced variation in nominal ground resolution cell size,
- scanning system one-dimensional relief displacement, and
- scanning system tangential scale distortion.

Image Offset (Skew) Caused by Earth Rotation Effects

Earth-observing Sun-synchronous satellites are normally in fixed orbits that collect a path (or swath) of imagery as the satellite makes its way from the north to the south in descending mode (Figure 7-1a). Meanwhile, the Earth below rotates on its axis from west to east making one complete revolution every 24 hours. This interaction between the fixed orbital path of the remote sensing system and the Earth's rotation on its axis *skews* the geometry of the imagery collected. For example, consider just three hypothetical scans of 16 lines each obtained by the Landsat Enhanced Thematic Mapper Plus (ETM^+). If the data are not deskewed, they will appear in the dataset incorrectly, as shown in Figure 7-1b. While this looks correct, it is not. This matrix does not take into account the Earth rotation

effects. The data in this dataset are actually skewed to the *east* by a predictable amount.

Conversely, if the remotely sensed data are deskewed, then all of the pixels associated with a single scan (containing 16 lines of data) will be offset (adjusted) by the digital image processing system a certain amount to the *west* (Figure 7-1c). *Deskewing* is defined as the systematic displacement of pixels westward in a frame of imagery to correct for the interaction of the satellite sensor system's angular velocity and the Earth's surface velocity. This adjustment places all of the pixels in the scan in their proper positions relative to adjacent scans. Null values are added to the dataset to maintain the integrity of the raster (matrix) format. The amount of shift or displacement to the west is a function of the relative velocities of both the satellite and the Earth and the length of the image frame that is recorded.

Most satellite image data providers automatically deskew the data that is delivered to customers using the logic shown in Figure 7-1c. First, the surface velocity of the Earth, v_{earth}, is computed,

$$v_{earth} = \omega_{earth}\, r \cos\lambda \tag{7.1}$$

where r is the radius of the Earth (6.37816 Mm) and ω_{earth} is the Earth's rotational velocity (72.72 μrad s^{-1}) at a specific latitude, λ. Therefore, for Charleston, SC, located at 33° N latitude, the surface velocity of the Earth is:

$$v_{earth} = 72.72\ \mu\text{rad s}^{-1} \times 6.37816\ \text{Mm} \times 0.83867$$

$$v_{earth} = 389\ \text{m s}^{-1}.$$

Next we must determine the length of time it takes for the satellite to scan a typical remote sensing frame (F) of data on the ground. In this example we will consider a 185-km frame obtained by Landsats 1, 2, and 3 associated with the Multispectral Scanner and Landsats 4, 5, and 7 associated primarily with the Thematic Mapper and Enhanced Thematic Mapper Plus. Landsats 1, 2, and 3 have an angular velocity, $\omega_{land123}$, of 1.014 mrad s^{-1}. Landsat satellites 4, 5, and 7 have an angular velocity, $\omega_{land457}$, of 1.059 mrad s^{-1} (Williams, 2003).[1] Therefore, a typical 185-km frame of Landsat MSS imagery (WRS1 orbit) would be scanned in:

$$s_t = \frac{L}{r \times \omega_{land123}} \tag{7.2}$$

$$s_t = \frac{185\ \text{km}}{(6.37816\ \text{Mm})(1.014\ \text{mrad s}^{-1})} = 28.6\ s.$$

A 185-km frame of Landsat 4, 5 and 7 imagery (WRS2 orbit) would be scanned in:

$$s_t = \frac{185\ \text{km}}{(6.37816\ \text{Mm})(1.059\ \text{mrad s}^{-1})} = 27.4\ s.$$

Therefore, during the time that the 185-km frame of Landsat 1, 2, or MSS imagery was collected at Charleston, SC (32° N latitude), the surface of the Earth moved to the east by:

$$\Delta x_{east} = v_{earth} \times s_t \tag{7.3}$$

$$\Delta x_{east} = 389\ \text{m s}^{-1} \times 28.6\ \text{s} = 11.12\ \text{km}.$$

Similarly, during the time a 185-km frame of Landsat ETM^+ imagery is collected, the surface of the Earth moves to the east by:

$$\Delta x_{east} = 389\ \text{m s}^{-1} \times 27.39\ \text{s} = 10.65\ \text{km}.$$

This is approximately 6% of the 185-km frame size (e.g., 11.12/185 = 0.06; 10.65/185 = 0.057) for Landsats 1, 2, and 3 MSS and Landsat Thematic Mappers 4, 5, and 7 data. Fortunately, much of the satellite imagery provided by commercial and public remote sensing data providers have already been deskewed by the appropriate amount per line scan.

The skewing described holds true for line-scan sensors such as Landsat TM and ETM^+ and for pushbroom sensors such as SPOT HRV, DigitalGlobe's WorldView-2, and GeoEye's GeoEye-1. Every fixed-orbit remote sensing system that collects data while the Earth rotates on its axis will contain frames of imagery that are skewed.

Scanning System-Induced Variation in Ground Resolution Cell Size

A large amount of remote sensor data is acquired using scanning systems (e.g., Landsat 7, ASTER). Fortunately, an orbital multispectral scanning system scans through just a few degrees off-nadir as it collects data hundreds of kilometers above the Earth's surface (e.g., Landsat 7 data are collected at 705 km AGL). This configuration minimizes the amount of distortion introduced by the scanning system. Conversely, a suborbital multispectral scanning system may be operating just tens of kilometers AGL with a scan field of view of perhaps 70 degrees. This introduces numerous types of geometric distortion that can be difficult to correct.

1. For Landsats 1, 2, and 3 (WRS1 orbit) the angular velocity is (251 paths/cycle × 2π × 1,000 mrad/path) / (18 days/cycle × 86,400 s/day) = 1.014 mrad/s. For Landsats 4, 5, and 7 (WRS2 orbit) the angular velocity is (233 paths/cycle × 2 π × 1,000 mrad/path) / (16 days/cycle × 86,400 s/day) = 1.059 mrad/s.

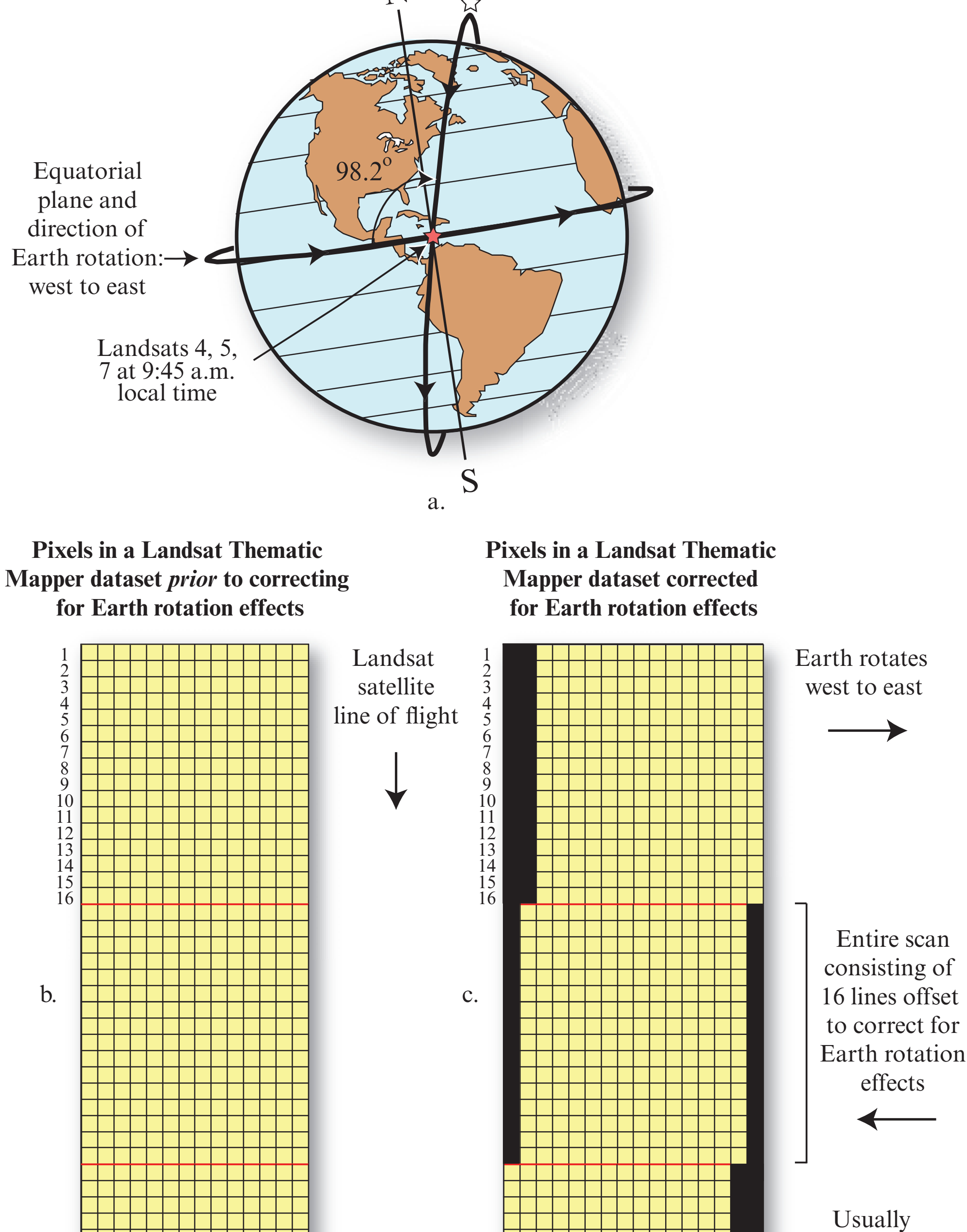


FIGURE 7–1 a) Landsat satellites 4, 5, and 7 are in a Sun-synchronous orbit with an angle of inclination of 98.2°. The Earth rotates on its axis from west to east as imagery is collected. b) Pixels in three hypothetical scans (consisting of 16 lines each) of Landsat TM data. While the matrix (raster) may look correct, it actually contains systematic geometric distortion caused by the angular velocity of the satellite in its descending orbital path in conjunction with the surface velocity of the Earth as it rotates on its axis while collecting a frame of imagery. c) The result of adjusting (deskewing) the original Landsat TM data to the west to compensate for Earth rotation effects. Landsats 4, 5, and 7 use a bidirectional cross-track scanning mirror (NASA, 1998).

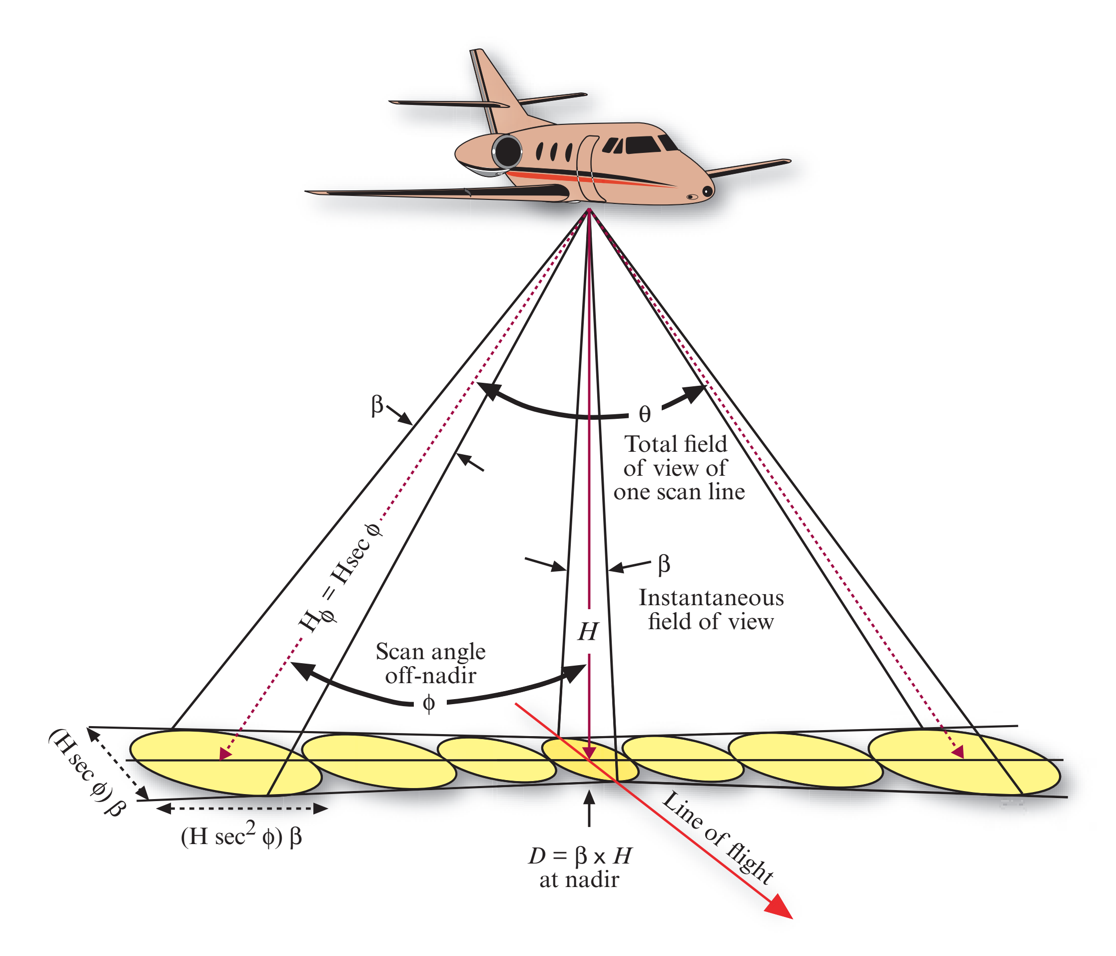


FIGURE 7–2 The ground resolution cell size along a single across-track scan is a function of a) the distance from the aircraft to the observation where H is the altitude of the aircraft above ground level (AGL) at nadir and $H \sec \phi$ off-nadir; b) the instantaneous-field-of-view of the sensor, β, measured in radians; and c) the scan angle off-nadir, ϕ. Pixels off-nadir have semimajor and semiminor axes (diameters) that define the resolution cell size. The total field of view of one scan line is θ. One-dimensional relief displacement and tangential scale distortion occur in the direction perpendicular to the line of flight and parallel with a line scan.

The ground swath width (*gsw*) is the length of the terrain strip remotely sensed by the system during one complete across-track sweep of the scanning mirror. It is a function of the total angular field of view of the sensor system, θ, and the altitude of the sensor system above ground level, H as shown in Figure 7-2. It is computed using the equation:

$$gsw = \tan\left(\frac{\theta}{2}\right) \times H \times 2\,. \tag{7.4}$$

For example, the ground swath width of an across-track scanning system with a 90° total field of view and an altitude above ground level of 6,000 m would be 12,000 m:

$$gsw = \tan\left(\frac{90}{2}\right) \times 6{,}000 \times 2$$

$$gsw = 1 \times 6{,}000 \times 2$$

$$gsw = 12{,}000 \text{ m.}$$

Most scientists using across-track scanner data use only approximately the central 70% of the swath width (35% on each side of nadir) primarily because ground resolution elements have larger cell sizes the farther they are away from nadir.

The diameter of the circular ground area viewed by the sensor, D, at nadir is a function of the instantaneous-field-of-view, β, of the scanner measured in milliradians (mrad) and the altitude of the scanner above ground level, H, where $D = \beta \times H$. As the scanner's instantaneous-field-of-view moves away from nadir on either side, the circle becomes an ellipsoid because the distance from the aircraft to the resolution cell is increasing, as shown in Figure 7-2. In fact, the distance from the aircraft to the resolution cell on the ground, H_ϕ, is a function of the scan angle off-nadir, ϕ, at the time of data collection and the true altitude of the aircraft, H (Lillesand et al., 2008):

$$H_\phi = H \times \sec\phi\,. \tag{7.5}$$

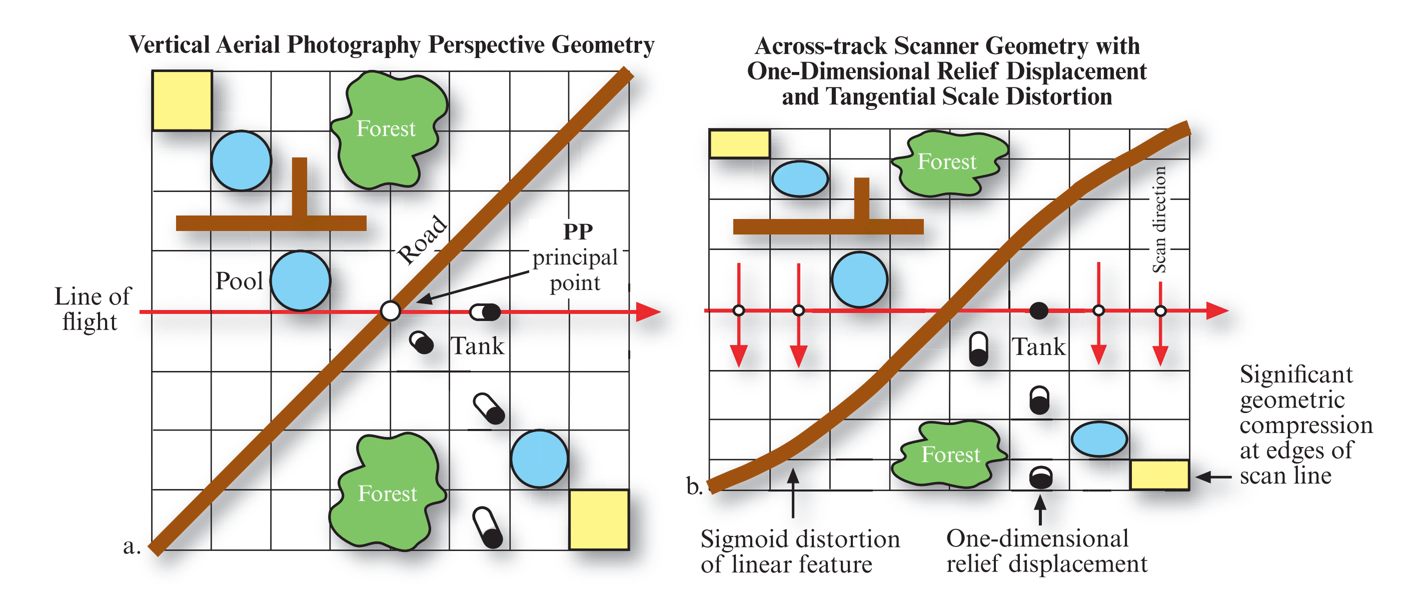


FIGURE 7–3 a) Hypothetical perspective geometry of a vertical aerial photograph obtained over level terrain. Four 50-ft-tall tanks are distributed throughout the landscape and experience varying degrees of radial relief displacement the farther they are from the principal point (PP). b) Across-track scanning system introduces one-dimensional relief displacement perpendicular to the line of flight and tangential scale distortion and compression the farther the object is from nadir. Linear features trending across the terrain are often recorded with *s*-shaped or sigmoid curvature characteristics due to tangential scale distortion and image compression.

Thus, the size of the ground-resolution cell increases as the angle increases away from nadir. The nominal (average) diameter of the elliptical resolution cell, D_ϕ, at this angular location from nadir has the dimension:

$$D_\phi = (H \times \sec\phi) \times \beta \quad (7.6)$$

in the direction of the line of flight, and

$$D_\phi = (H \times \sec^2\phi) \times \beta \quad (7.7)$$

in the orthogonal (perpendicular) scanning direction.

Scientists using across-track scanner data usually concern themselves only with the spatial ground resolution of the cell at nadir, *D*. If it is necessary to perform precise quantitative work on pixels some angle ϕ off-nadir, then it may be important to remember that the radiant flux recorded is an integration of the radiant flux from all the surface materials in a ground-resolution cell with a constantly changing diameter. Using only the central 70% of the swath width reduces the impact of the larger pixels found at the extreme edges of the swath.

Scanning System One-Dimensional Relief Displacement

Truly vertical aerial photographs have a single principal point directly beneath the aircraft at nadir at the instant of exposure. This perspective geometry causes all objects that rise above the local terrain elevation to be displaced from their proper planimetric position radially outward from the principal point. For example, the four hypothetical tanks in Figure 7-3a are each 50 ft. high. The greater the distance from the principal point, the greater the radial relief displacement of the top of the tank away from its base.

Images acquired using an across-track scanning system also contain relief displacement. However, instead of being radial from a single principal point, the displacement takes place in a direction that is perpendicular to the line of flight for each and every scan line, as shown in Figure 7-3b. In effect, the ground-resolution element at nadir functions like a principal point for each scan line. At nadir, the scanning system looks directly down on the tank, and it appears as a perfect circle in Figure 7-3b. The greater the height of the object above the local terrain and the greater the distance of the top of the object from nadir (i.e., the line of flight), the greater the amount of *one-dimensional relief displacement* present. One-dimensional relief displacement is introduced in both directions away from nadir for each sweep of the across-track mirror.

Although some aspects of one-dimensional relief displacement may be of utility for visual image interpretation, it seriously displaces the tops of objects projecting above the local terrain from their true planimetric position. Maps produced from such imagery contain serious planimetric errors.

Scanning System Tangential Scale Distortion

The mirror on an across-track scanning system rotates at a constant speed and typically views from 70º to 120º of terrain during a complete line scan. Of course, the amount depends on the specific sensor system. From Figure 7-2 it is clear that the terrain directly beneath the aircraft (at nadir) is closer to the aircraft than the terrain at the edges during a single sweep of the mirror. Therefore, because the mirror rotates at a constant rate, the sensor scans a shorter geographic distance at nadir than it does at the edge of the image. This relationship tends to *compress* features along an axis that is perpendicular to the line of flight. The greater the distance of the ground-resolution cell from nadir, the greater the image scale compression. This is called *tangential scale distortion.* Objects near nadir exhibit their proper shape. Objects near the edge of the flightline become compressed and their shape distorted. For example, consider the tangential geometric distortion and compression of the circular swimming pools and one hectare of land the farther they are from nadir in the hypothetical diagram (Figure 7-3b).

This tangential scale distortion and compression in the far range also causes linear features such as roads, railroads, utility rights-of-way etc., to have an *s-shape* or *sigmoid distortion* when recorded on scanner imagery (Figure 7-3b). Interestingly, if the linear feature is parallel with or perpendicular to the line of flight, it does not experience sigmoid distortion.

Even single flightlines of aircraft MSS data are difficult to rectify to a standard map projection because of aircraft roll, pitch, and yaw during data collection (van der Meer et al., 2009). Notches in the edge of a flightline of data are indicative of aircraft roll. Such data require significant human and machine resources to make the data planimetrically accurate. Most commercial data providers now place GPS on the aircraft to obtain precise flightline coordinates, which are useful when rectifying the aircraft MSS data.

External Geometric Errors

External geometric errors are usually introduced by phenomena that vary in nature through space and time. The most important external variables that can cause geometric error in remote sensor data are random movements by the aircraft (or spacecraft) at the exact time of data collection, which usually involve:

- altitude changes, and/or
- attitude changes (roll, pitch, and yaw).

Altitude Changes

A remote sensing system is ideally flown at a constant altitude above ground level (AGL) resulting in imagery with a uniform scale all along the flightline. For example, a frame camera with a 12-in. focal length lens flown at 20,000 ft. AGL will yield 1:20,000-scale imagery. If the aircraft or spacecraft gradually changes its altitude along a flightline, then the scale of the imagery will change (Figure 7-4a). Increasing the altitude will result in smaller-scale imagery (e.g., 1:25,000-scale). Decreasing the altitude of the sensor system will result in larger-scale imagery (e.g., 1:15,000). The same relationship holds true for digital remote sensing systems that collect imagery on a pixel by pixel basis. The diameter of the spot size on the ground (*D*; the nominal spatial resolution) is a function of the instantaneous-field-of-view (β) and the altitude above ground level (*H*) of the sensor system, i.e., $D = \beta \times H$.

It is important to remember, however, that scale changes can be introduced into the imagery even when the remote sensing system is flown at a constant elevation above ground level. This occurs when the terrain gradually increases or decreases in elevation (i.e., it moves closer to or farther away from the sensor system). For example, if the terrain surface is at 1,000 ft. AGL at the start of a flightline and 2,000 ft. AGL at the end of the flightline, then the scale of the imagery will become larger as the flightline progresses. Remote sensing platforms do not generally attempt to adjust for such gradual changes in elevation. Rather, it is acknowledged that scale changes will exist in the imagery and that the use of geometric rectification algorithms will normally be used to minimize the effects. The methods for adjusting for changes in scale using ground control points and geometric rectification coefficients will be discussed shortly.

Attitude Changes

Satellite platforms are usually stable because they are not buffeted by atmospheric turbulence or wind. Conversely, aircraft flying at suborbital altitudes must continuously contend with atmospheric downdrafts, updrafts, head-winds, tail-winds, and cross-winds when collecting remote sensor data. Even when the remote sensing platform maintains a constant altitude AGL, it may rotate randomly about three separate axes that are commonly referred to as *roll*, *pitch*, and *yaw* (Figure 7-4b). For example, sometimes the fuselage remains horizontal, but the aircraft rolls from side to side about the *x*-axis (direction of flight) some $\omega°$, introducing compression and/or expansion of the imagery in the near- and far-ranges perpendicular to the line of flight (Wolf et al., 2013). Similarly, the aircraft may be flying in the intended direction but the nose pitches up or down a certain $\phi°$ about the *y*-axis. If the nose pitches down, the imagery will be compressed in the fore-direction (toward the nose of the aircraft) and expanded in the aft-direction (toward the tail). If the nose pitches up, the imagery will be compressed in the aft-direction and

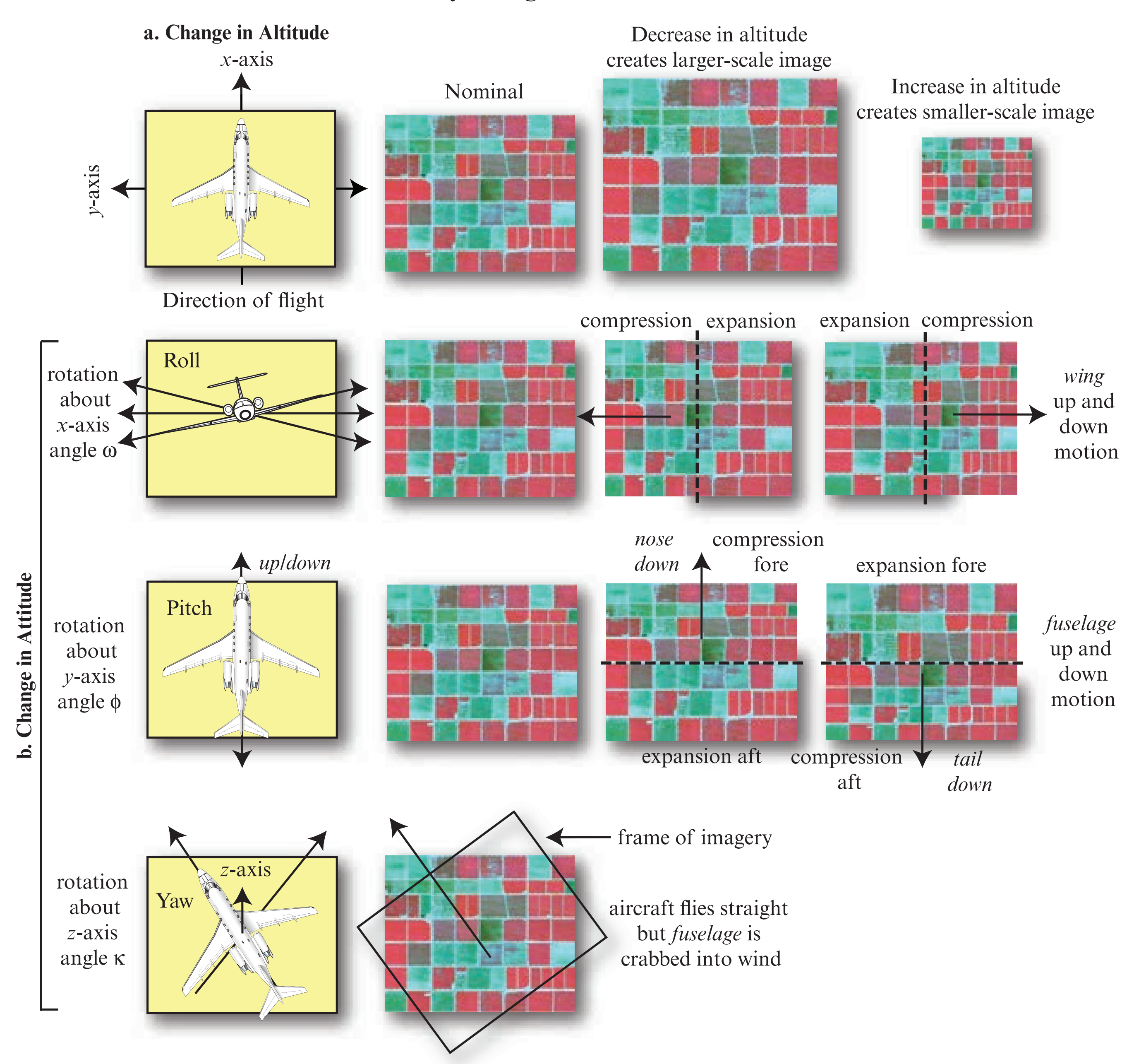


FIGURE 7–4 a) Geometric modification in imagery may be introduced by changes in the aircraft or satellite platform *altitude* above ground level (AGL) at the time of data collection. Increasing altitude results in smaller-scale imagery while decreasing altitude results in larger-scale imagery. b) Geometric modification may also be introduced by aircraft or spacecraft changes in *attitude*, including roll, pitch, and yaw. An aircraft flies in the *x*-direction. Roll occurs when the aircraft or spacecraft fuselage maintains directional stability but the wings move up or down, i.e. they rotate about the *x*-axis angle (omega: ω). Aircraft pitch occurs when the wings are stable but the fuselage nose or tail moves up or down, i.e., they rotate about the *y*-axis angle (phi: ϕ). Yaw occurs when the wings remain parallel but the fuselage is forced by wind to be oriented some angle to the left or right of the intended line of flight, i.e., it rotates about the z-axis angle (kappa: κ). Thus, the plane flies straight but all remote sensor data are displaced by κ°. Remote sensing data often are distorted due to a combination of changes in *altitude* and *attitude* (roll, pitch, and yaw).

expanded in the fore-direction. Occasionally a remote sensing platform experiences significant headwinds (or tailwinds) that must be compensated for in order to fly in a straight direction. When this occurs, it is possible for the pilot to “crab” (i.e., angle) the aircraft fuselage into the wind $\kappa°$ about the z-axis. The result is an accurate flightline but imagery that is oriented some $\kappa°$ from the intended flightline (Figure 7-4b).

High-quality satellite and aircraft remote sensing systems often have gyro-stabilization equipment that, in effect, isolates the sensor system from the roll and pitch movements of the aircraft. Remote sensing systems without stabilization equipment introduce some geometric error into the remote sensing dataset through variations in roll, pitch, and yaw that can only be corrected using ground control points.

Ground Control Points

Geometric distortions introduced by sensor system attitude (roll, pitch, and yaw) and/or altitude changes can be corrected using ground control points and appropriate mathematical models (e.g., Im et al., 2009). A **ground control point** (GCP) is a location on the surface of the Earth (e.g., a road intersection) that can be identified on the imagery and located accurately on a map. The image analyst must be able to obtain two distinct sets of coordinates associated with each GCP:

- image coordinates specified in i rows and j columns, and
- map coordinates (e.g., x, y measured in degrees of latitude and longitude, feet in a state plane coordinate system, or meters in a Universal Transverse Mercator projection).

The paired coordinates (i, j and x, y) from many GCPs (e.g., 20) can be modeled to derive geometric transformation coefficients (Wolf et al., 2013). These coefficients may then be used to geometrically rectify the remotely sensed data to a standard datum and map projection.



Types of Geometric Correction

Most commercially available remote sensor data (e.g., from SPOT Image Inc., DigitalGlobe Inc., GeoEye, Inc.) already have much of the systematic geometric error removed. Unless otherwise processed, however, the unsystematic random error remains in the image, making it nonplanimetric (i.e., the pixels are not in their correct x, y planimetric map position). This section focuses on two common geometric correction procedures often used by scientists to make the digital remote sensor data of value:

- image-to-map rectification, and
- image-to-image registration.

The general rule of thumb is to rectify remotely sensed data to a standard map projection whereby it may be used in conjunction with other spatial information in a GIS to solve problems (Merchant and Narumalani, 2009; Purkis and Klemas, 2011). Therefore, most of the discussion will focus on image-to-map rectification.

Image-to-Map Rectification

Image-to-map rectification is the process by which the geometry of an image is made planimetric. Whenever accurate area, direction, and distance measurements are required, image-to-map geometric rectification should be performed. It may not, however, remove all the distortion caused by topographic relief displacement in images. The image-to-map rectification process normally involves selecting GCP image pixel coordinates (row and column) with their map coordinate counterparts (e.g., meters northing and easting in a Universal Transverse Mercator map projection). For example, Figure 7-5 displays three GCPs (13, 14, and 16) easily identifiable by an image analyst in both the U.S. Geological Survey 1:24,000-scale 7.5-minute quadrangle and an unrectified Landsat TM band 4 image of Charleston, SC. It will be demonstrated how the mathematical relationship between the image coordinates and map coordinates of the selected GCPs is computed and how the image is made to fit the geometry of the map.

In the U.S., there are several alternatives to obtaining accurate GCP *map* coordinate information for image-to-map rectification, including:

- hard-copy planimetric maps (e.g., U.S. Geological Survey 7.5-minute 1:24,000-scale topographic maps) where GCP coordinates are extracted using simple ruler measurements or a coordinate digitizer;
- digital planimetric maps (e.g., the U.S. Geological Survey’s digital 7.5-minute topographic map series) where GCP coordinates are extracted directly from the digital map on the screen;
- digital orthophotoquads that are already geometrically rectified (e.g., U.S. Geological Survey digital orthophoto quarter quadrangles —DOQQ); and/or
- global positioning system (GPS) instruments that may be taken into the field to obtain the coordinates of objects to within ±20 cm if the GPS data are differentially corrected (e.g., Jensen and Jensen, 2013).

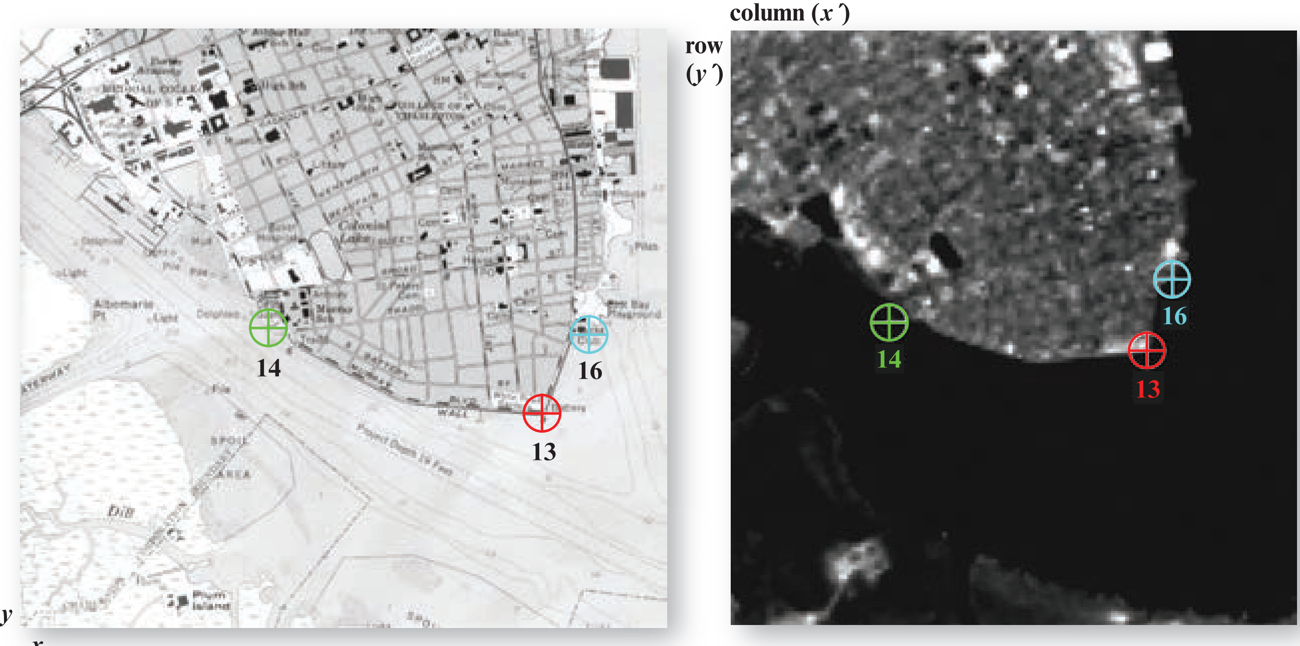


FIGURE 7–5 Example of *image-to-map rectification*. a) U.S. Geological Survey 7.5-minute 1:24,000-scale topographic map of Charleston, SC, with three ground control points identified (13, 14, and 16). The GCP map coordinates are measured in meters easting (*x*) and northing (*y*) in a Universal Transverse Mercator projection. b) Unrectified 11/09/82 Landsat TM band 4 image with the three ground control points identified. The image GCP coordinates are measured in rows and columns. Image courtesy of NASA.

GPS collection of map coordinate information to be used for image rectification is especially effective in poorly mapped regions of the world or where rapid change has made existing maps obsolete (Jensen et al., 2002).

Image-to-Image Registration

Image-to-image registration is the translation and rotation alignment process by which two images of like geometry and of the same geographic area are positioned coincident with respect to one another so that corresponding elements of the same ground area appear in the same place on the registered images. This type of geometric correction is used when it is *not* necessary to have each pixel assigned a unique *x, y* coordinate in a map projection. For example, we might want to make a cursory examination of two images obtained on different dates to see if any change has taken place. While it is possible to rectify both of the images to a standard map projection and then evaluate them (image-to-map rectification), this may not be necessary to simply identify the change that has taken place between the two images.

Hybrid Approach to Image Rectification/Registration

It is interesting that the same general image processing principles are used in both image rectification and image registration. The difference is that in image-to-map rectification the reference is a map in a standard map projection, while in image-to-image registration the reference is another image. It should be obvious that if a rectified image is used as the reference base (rather than a traditional map) any image registered to it will inherit the geometric errors existing in the reference image. Because of this characteristic, most serious Earth science remote sensing research is based on analysis of data that have been rectified to a map base. However, when conducting rigorous change detection between two or more dates of remotely sensed data, it may be useful to select a *hybrid* approach involving both image-to-map rectification and image-to-image registration (Jensen et al., 1993).

An example of the hybrid approach is demonstrated in Figure 7-6 where an October 14, 1987, Landsat TM image is being registered to a rectified November 9, 1982, Landsat TM scene. In this case, the 1982 base year image was previously rectified to a Universal

Transverse Mercator map projection with 30×30 m pixels. Ground control points are being selected to register the 1987 image to the rectified 1982 base year image. It is often very difficult to locate good ground control points in remotely sensed data, especially in rural areas (e.g., forests, wetland, and water bodies). The use of the *rectified* base year image as the map allows many more common GCPs to be located in the unrectified 1987 imagery. For example, edges of water bodies and fields or the intersection of small stream segments are not usually found on a map but may be easy to identify in the rectified and unrectified imagery.

The optimum method of selecting the ground control points is to have both the rectified base year image (or reference map) and the image to be rectified on the screen at the same time (Figure 7-6). This dual display greatly simplifies GCP selection. Some image processing systems even allow the GCP selected to be reprojected onto the image to be corrected (with the appropriate transformation coefficients, to be discussed shortly) to determine the quality of the GCP point. Also, some systems allow the analyst to extract floating point row and column coordinates of GCPs (instead of just integer values) through the use of a chip extraction algorithm that zooms in and does subpixel sampling, as demonstrated in Figure 7-6. GCP subpixel row and column coordinates often improve the precision of the image-to-map rectification or image-to-image registration. Some scientists have developed analytical methods of automatically extracting GCPs common to two images that can be used during image-to-image registration. However, most image-to-map rectification still relies heavily on human selection of GCPs.

The following example focuses on image-to-map geometric rectification because it is the most frequently used method of removing geometric distortion from remotely sensed data.

Image-to-Map Geometric Rectification Logic

Two basic operations must be performed to geometrically rectify a remotely sensed image to a map coordinate system:

1. The geometric relationship between the input pixel coordinates (column and row; referred to as x', y') and the associated map coordinates of this same point (x, y) must be identified (Figures 7-5 and 7-6). A number of GCP pairs are used to establish the nature of the geometric coordinate transformation that must be applied to rectify or fill every pixel in the output image (x, y) with a value from a pixel in the unrectified input image (x', y'). This process is called *spatial interpolation*.
2. Pixel brightness values must be determined. Unfortunately, there is no direct one-to-one relationship between the movement of input pixel values to output pixel locations. It will be shown that a pixel in the rectified output image often requires a value from the input pixel grid that does not fall neatly on a row-and-column coordinate. When this occurs, there must be some mechanism for determining the brightness value (BV) to be assigned to the output rectified pixel. This process is called *intensity interpolation.*

Spatial Interpolation Using Coordinate Transformations

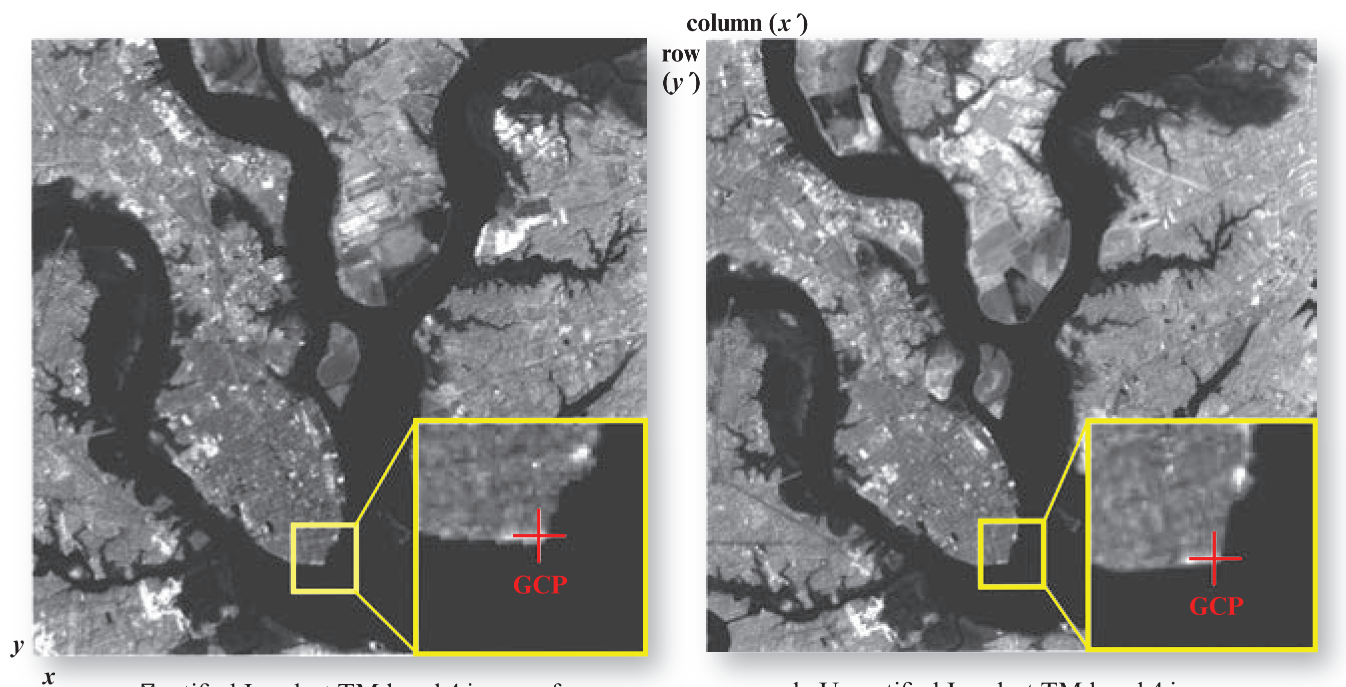
As discussed earlier, some distortions in remotely sensed data may be removed or mitigated using techniques that model systematic orbital and sensor characteristics. Unfortunately, this does not remove error produced by changes in attitude (roll, pitch, and yaw) or altitude. Such errors are generally unsystematic and are often removed by identifying GCPs in the original imagery and on the reference map and then mathematically modeling the geometric distortion. Image-to-map rectification requires that polynomial equations be fit to the GCP data using least-squares criteria to model the corrections directly in the image domain without explicitly identifying the source of the distortion (Novak, 1992; Bossler, 2010). Depending on the distortion in the imagery, the number of GCPs used, and the degree of topographic relief displacement in the area, higher-order polynomial equations may be required to geometrically correct the data. The *order* of the rectification is simply the highest exponent used in the polynomial. For example, Figure 7-7 demonstrates how different-order transformations fit a hypothetical surface. Generally, for moderate distortions in a relatively small area of an image (e.g., a quarter of a Landsat TM scene), a first-order, six-parameter, affine (*linear*) transformation is sufficient to rectify the imagery to a geographic frame of reference.

This type of transformation can model six kinds of distortion in the remote sensor data, including (Novak, 1992; Buiten and Van Putten, 1997):

- translation in x and y,
- scale changes in x and y,
- skew, and
- rotation.

Input-to-Output (Forward) Mapping: When all six of the operations are combined into a single expression it becomes:

Image-to-Image Hybrid Registration



a. Rectified Landsat TM band 4 image of Charleston, SC, obtained on November 9, 1982.

b. Unrectified Landsat TM band 4 image obtained on October 14, 1987.

FIGURE 7–6 Example of image-to-image *hybrid* registration. a) Previously rectified Landsat TM band 4 data obtained on November 9, 1982, resampled to 30 × 30 m pixels using nearest-neighbor resampling logic and a UTM map projection. b) Unrectified October 14, 1987, Landsat TM band 4 data to be registered to the rectified 1982 Landsat scene. Original imagery courtesy of NASA.

$$x = a_0 + a_1x' + a_2y' \tag{7.8}$$
$$y = b_0 + b_1x' + b_2y',$$

where x and y are positions in the *output*-rectified image or map, and x' and y' represent corresponding positions in the original *input* image or map (Figure 7-8a). These two equations can be used to perform what is commonly referred to as *input-to-output* or *forward-mapping*. The equations function according to the logic shown in Figure 7-8a. In this example, each pixel in the *input* grid (e.g., value 15 at $x', y' = 2, 3$) is sent to an x,y location in the output image according to the six coefficients shown in Figure 7-8a.

This *forward mapping* logic works well if we are simply rectifying the location of discrete coordinates found along a linear feature such as a road in a vector map. In fact, cartographic mapping and geographic information systems typically rectify vector data using forward mapping logic. However, when we are trying to fill a rectified *output* grid (matrix) with values from an unrectified *input* image, forward mapping logic does not work well. The basic problem is that the six coefficients may require that value 15 from the x', y' location 2, 3 in the input image be located at a floating point location in the output image at x, y = 5, 3.5, as shown in Figure 7-8a. The output x, y location does not fall exactly on an integer x and y output map coordinate. *In fact, using forward mapping logic can result in output matrix pixels with* ***no*** *output value* (Wolberg, 1990). This is a serious condition and one that reduces the utility of the remote sensor data for useful applications. For this reason, most remotely sensed data are geometrically rectified using *output-to-input* or *inverse mapping* logic (Niblack, 1986; Richards, 2013).

Output-to-Input (Inverse) Mapping: The use of *output-to-input,* or *inverse mapping* logic, is based on the following two equations:

$$x' = a_0 + a_1x + a_2y \tag{7.9}$$
$$y' = b_0 + b_1x + b_2y$$

where x and y are positions in the *output*-rectified image or map, and x' and y' represent corresponding positions in the original *input* image. If desired, this relationship can also be written in matrix notation:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_0 \\ b_0 \end{bmatrix}. \tag{7.10}$$

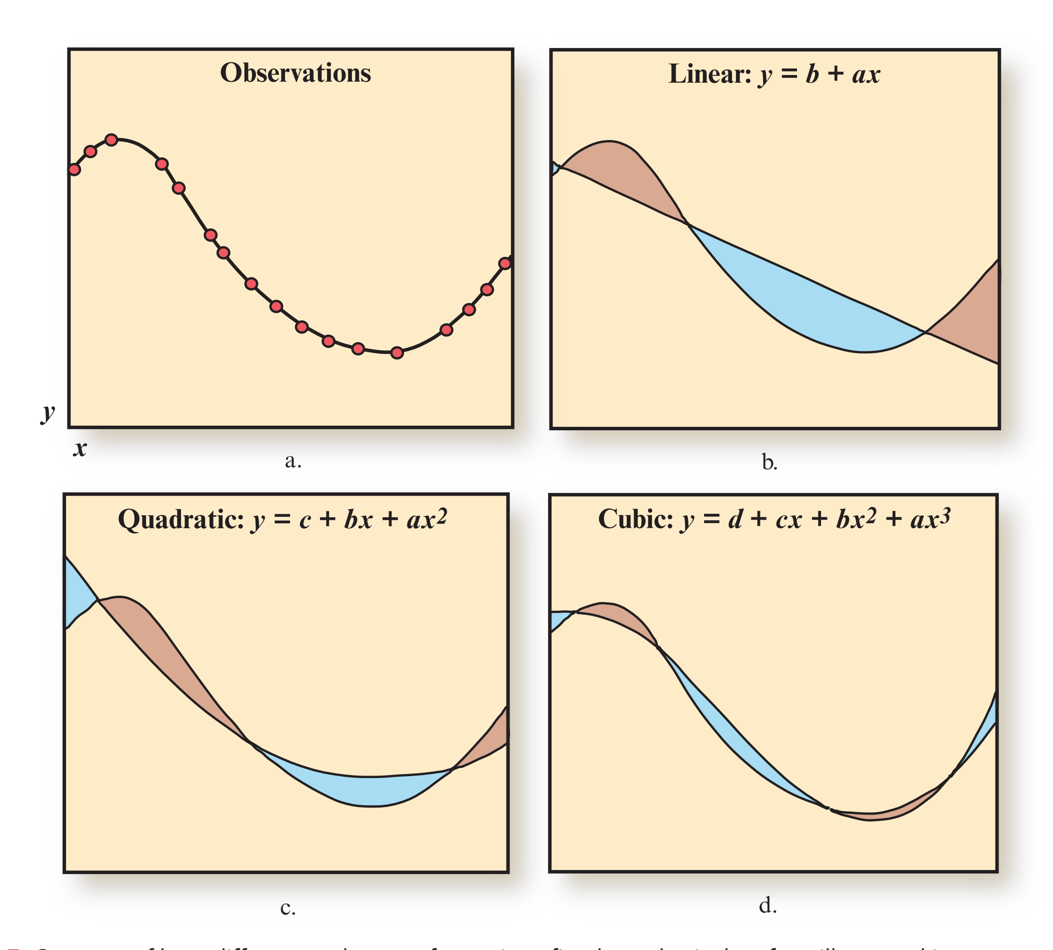


FIGURE 7–7 Concept of how different-order transformations fit a hypothetical surface illustrated in cross-section. a) The original observations. b) A first-order linear transformation fits a plane to the data. c) A second-order quadratic fit. d) A third-order cubic fit.

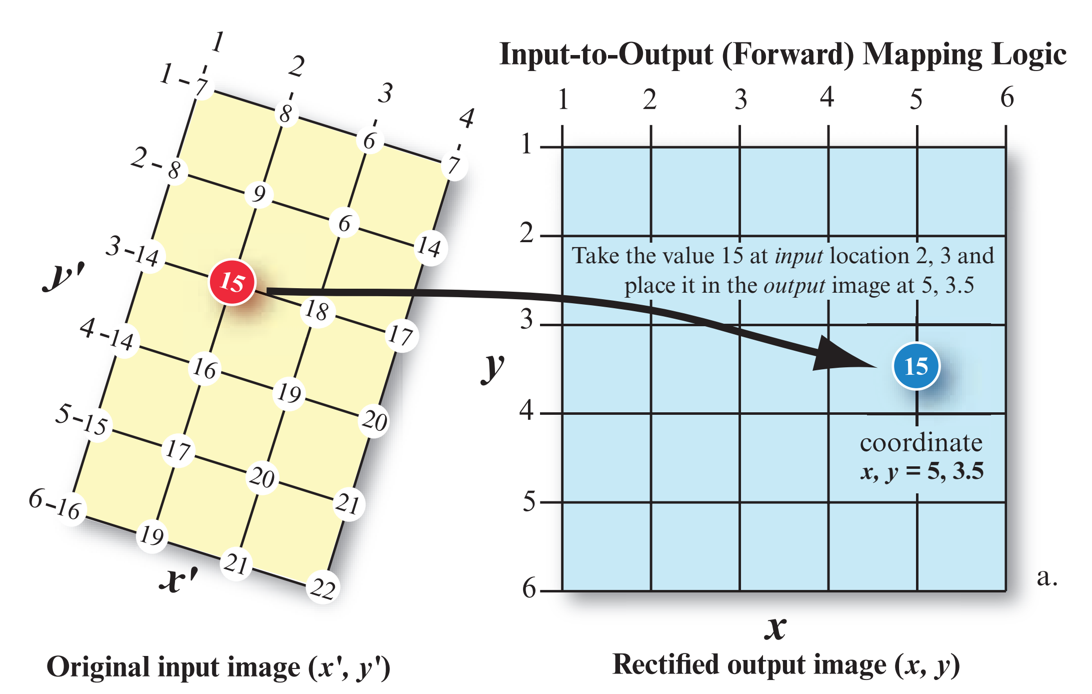
The rectified *output* matrix consisting of x (column) and y (row) coordinates is filled in the following systematic manner. Each output pixel location (e.g., x, y = 5, 4 in Figure 7-8b) is entered into Equation 7.9. The equation uses the six coefficients to determine where to go into the original *input* image to get a value (dashed line in Figure 7-8b). In this example, it requests a value from the floating point location x', y' = 2.4, 2.7. Interestingly, there is no value at this location. Nevertheless, it is possible to get a value (e.g., 15) if nearest-neighbor resampling logic is used (to be discussed shortly). The inverse mapping logic guarantees that there will be a value at every x, y coordinate (column and row) in the output image matrix. There will be no missing values. This procedure may seem backwards at first, but it is the only mapping function that avoids overlapping pixels and holes in the output image matrix (Schowengerdt, 2007; Konecny, 2014).

Sometimes it is difficult to gain an appreciation of exactly what impact the six coefficients in an affine (linear) transformation have on geographic coordinates. Therefore, consider Table 7-1, which demonstrates how the six coefficients (in parentheses) in a *forward* affine transformation influence the x, y coordinates of 16 pixels in a hypothetical 4 × 4 matrix. Note how *translation* (shifting) in x and y is controlled by the a_0 and b_0 coefficients, respectively. *Scale changes* in x and y are controlled by the a_1 and b_2 coefficients, respectively. *Shear/rotation* in x and y are controlled by coefficients a_2 and b_1, respectively. The three examples demonstrate how modifying the coefficients causes the pixel coordinates to be translated (shifted one unit in x and y), scaled (expanded by a factor of 2 in x and y), and sheared/rotated (Brown, 1992). Please remember that this particular example is based on *forward* mapping logic.

It is possible to use higher-order polynomial transformations to rectify remotely sensed data. For example, instead of using the six-parameter affine transformation previously discussed, we could use a second-order (quadratic) polynomial:

$$x' = c_0 + c_1x + c_2y + c_3xy + c_4x^2 + c_5y^2 \qquad (7.11)$$

$$y' = d_0 + d_1x + d_2y + d_3xy + d_4x^2 + d_5y^2 . \qquad (7.12)$$



$$x = a_0 + a_1 x' + a_2 y'$$
$$y = b_0 + b_1 x' + b_2 y' \quad \text{where}$$

x' and y' are locations in the original *input* image, and
x and y are locations in the rectified *output* image.

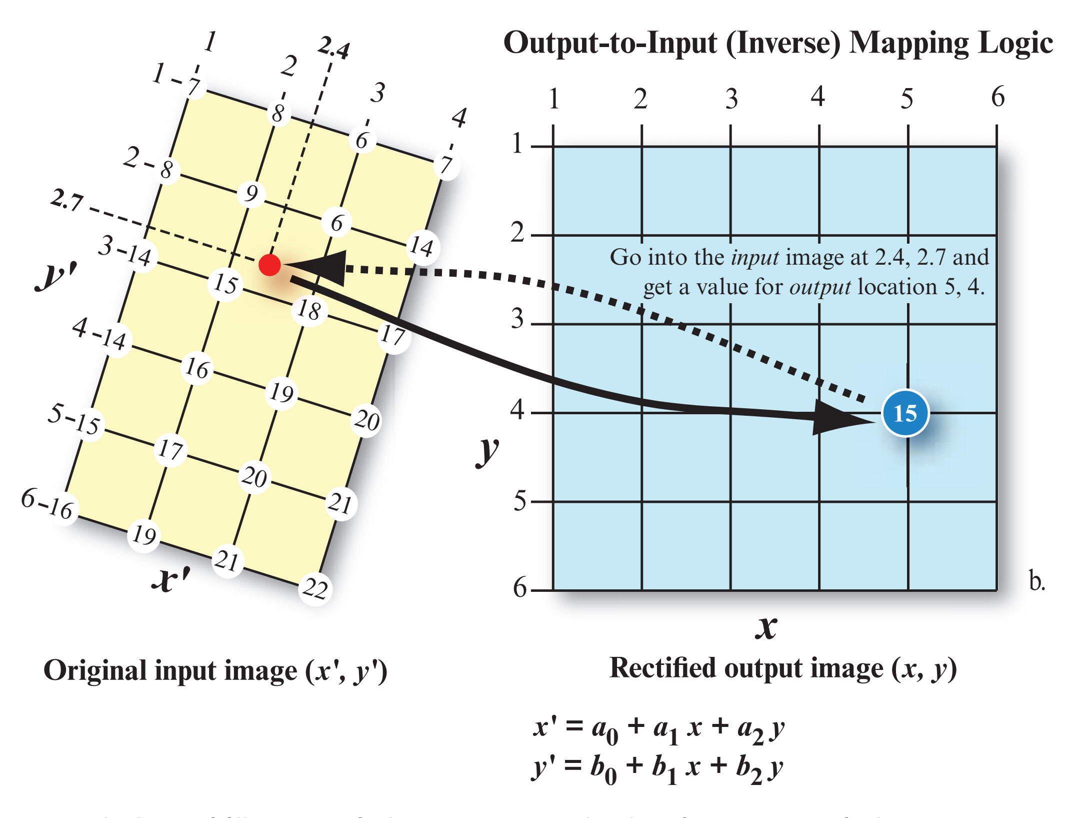


FIGURE 7–8 a) The logi c of filling a rectified out put matrix with values from an unrectified input image matrix using input-to-output (*forward*) mapping logic. b) The logic of filling a rectified output matrix wi th values from an unr ectified input image matrix using output-to-input (*inverse*) mapping logic and nearest-neighbor resampling. Output-to-input *inverse* mapping logic is the preferred methodology because it results in a rectified output matrix with values at every pixel location.

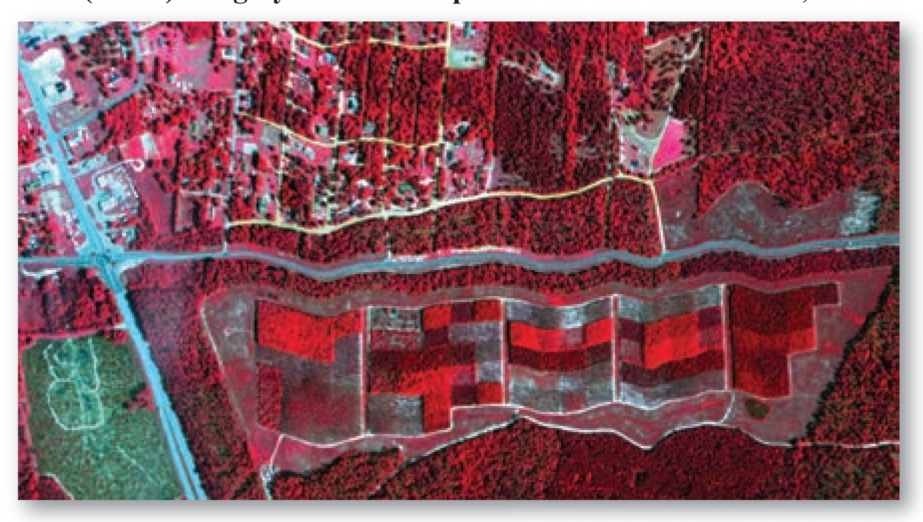
TABLE 7–1 *Forward* mapping example of how varying the six coefficients (in parentheses) in a linear (affine) transformation can impact the coordinates of 16 pixels in a simple 4 × 4 matrix. The equations are of the form $x_{predict} = a_0 + a_1x + a_2y$ and $y_{predict} = b_0 + b_1x + b_2y$. Translation (shfting) in x and y is controlled by the a_0 and b_0 coefficients, respectively. Scale changes in x and y are controlled by the a_1 and b_2 coefficients, respectively. Shear/rotation in x and y are controlled by coefficients a_2 and b_1, respectively.

Original Matrix $x_{predict} = (0) + (1)x + (0)y$ $y_{predict} = (0) + (0)x + (1)y$		Translation $x_{predict} = (1) + (1)x + (0)y$ $y_{predict} = (1) + (0)x + (1)y$		Scale $x_{predict} = (0) + (2)x + (0)y$ $y_{predict} = (0) + (0)x + (2)y$		Shear/Rotation $x_{predict} = (0) + (1)x + (2)y$ $y_{predict} = (0) + (2)x + (1)y$	
x = col	y = row	x	y	x	y	x	y
1	1	2	2	2	2	3	3
1	2	2	3	2	4	5	4
1	3	2	4	2	6	7	5
1	4	2	5	2	8	9	6
2	1	3	2	4	2	4	5
2	2	3	3	4	4	6	6
2	3	3	4	4	6	8	7
2	4	3	5	4	8	10	8
3	1	4	2	6	2	5	7
3	2	4	3	6	4	7	8
3	3	4	4	6	6	9	9
3	4	4	5	6	8	11	10
4	1	5	2	8	2	6	9
4	2	5	3	8	4	8	10
4	3	5	4	8	6	10	11
4	4	5	5	8	8	12	12

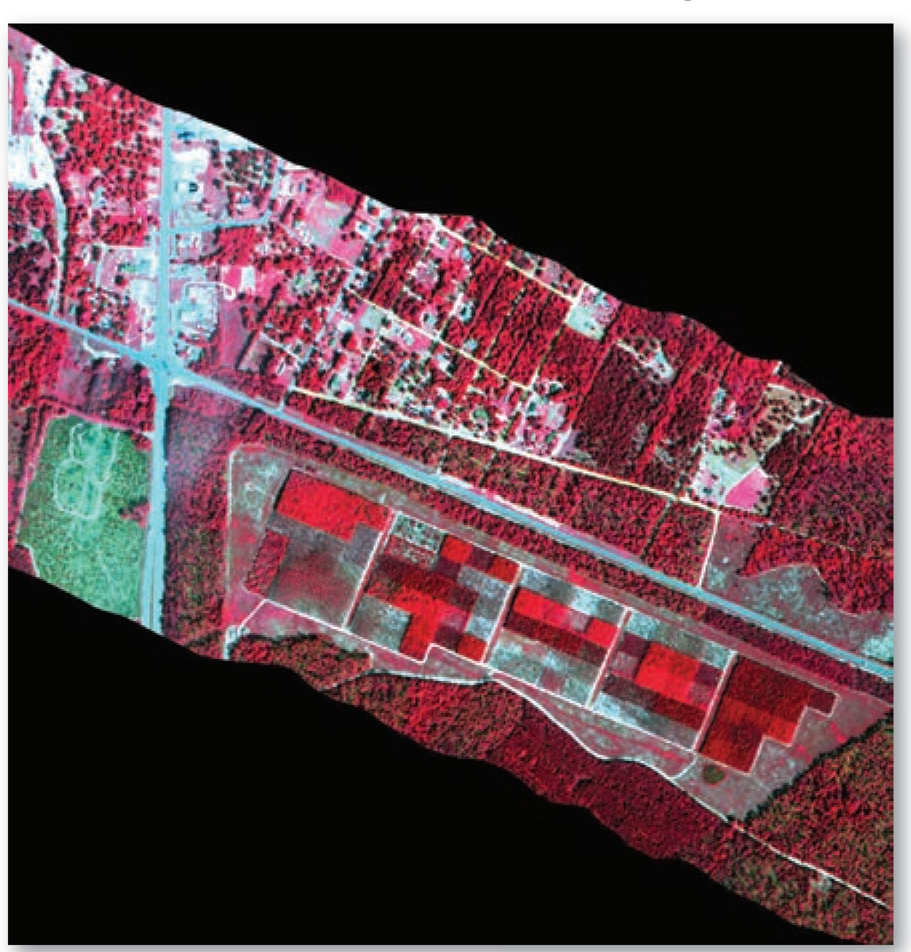
In theory, the higher the order of the polynomial, the more closely the coefficients should model the geometric error in the original (unrectified) input image (e.g., refer to Figure 7-7 to see how the various models fit one-dimensional data) and place the pixels in their correct planimetric positions in the rectified output matrix. Higher-order polynomials often produce a more accurate fit for areas immediately surrounding ground control points. However, other geometric errors may be introduced at large distances from the GCPs (Gibson and Power, 2000). In addition, the digital image processing system time required to geometrically rectify the remote sensor data using higher-order polynomials increases because of the greater number of mathematical operations that must be performed.

A general rule of thumb is to use a first-order affine polynomial whenever possible. Select a higher-order polynomial (e.g., second or third order) only when there are serious geometric errors in the dataset. These types of error are often found in imagery obtained from suborbital aerial platforms where roll, pitch, and yaw by the aircraft introduce unsystematic, nonlinear

Geometric Correction of Airborne Imaging Spectrometer for Applications (AISA) Imagery of Forest Experimental Plots near Aiken, SC



a. Uncorrected 1 × 1 m AISA 63-channel data collected on September 15, 2006.



b. Hyperspectral data geometrically corrected using GCPs, a third-order polynomial and nearest-neighbor resampling.

FIGURE 7–9 a) AISA 1 ⊠ 1 m hyperspectral data collected on September 15, 2006. Only three of the 63 channels are displayed (RGB = bands 760.8, 664.4, and 572.5 nm). b) The 63-channel dataset was rectified using a second-order polynomial to adjust for the significant geometric distortion in the original dataset caused by the turbulence during data collection. The forest experimental plots are subjected to various nutrient (fertilizer) and irrigation treatments (based on Im, J., Jensen, J. R., Coleman, M. and E. Nelson, 2009, "Hyperspectral Remote Sensing Analysis of Short Rotation Woody Crops Grown with Controlled Nutrient and Irrigation Treatments," *Geocarto International*, 24(4):293–312).

distortions. These distortions can only be modeled using a higher-order polynomial (e.g., Im et al., 2009). For example, the Airborne Imaging Spectrometer for Applications (AISA) hyperspectral imagery of forest experimental treatment plots shown in Figure 7-9 was geometrically rectified using a third-order polynomial. A first-order affine rectification was not sufficient to correct the significant warping in the imagery caused by the atmospheric turbulence during data collection. All pixels in this image are now in their appropriate geographic location (Im et al., 2009). Note that the roads in the rectified image are now straight.

Compute the Root-Mean-Squared Error of the Inverse Mapping Function Using the six coordinate transform coefficients that model distortions in the original scene, it is possible to use the output-to-input (inverse) mapping logic to transfer (relocate) pixel values from the original distorted image x', y' to the grid of the rectified output image, *x, y*. However, before applying the coefficients to create the rectified output image, it is important to determine how well the six coefficients derived from the least-squares regression of the initial GCPs account for the geometric distortion in the input image. The method used most often involves the computation of the root-mean-square error (RMS_{error}) for each of the ground control points (Wolf et al., 2013).

Let us consider for a moment the nature of the GCP data. We first identify a point in the image such as a road intersection. Its column and row coordinates in the original input image we will call x_{orig} and y_{orig}. The *x* and *y* position of the same road intersection is then measured from the reference map in degrees, feet, or meters. These two sets of GCP coordinates and many others selected by the analyst are used to compute the six coefficients discussed in Equation 7.9. Now, if we were to put the *map* *x* and *y* values for the first GCP back into Equation 7.9, with all the coefficients in place, we would get computed x' and y' values that are supposed to be the location of this point in the input image space. Ideally, x' would equal x_{orig} and y' would equal y_{orig}. Unfortunately, this is rarely the case. Any discrepancy between the values represents image geometric distortion not corrected by the six-coefficient coordinate transformation.

A simple way to measure such distortion is to compute the RMS_{error} for each ground control point by using the equation:

$$RMS_{error} = \sqrt{(x' - x_{orig})^2 + (y' - y_{orig})^2}, \qquad (7.13)$$

where x_{orig} and y_{orig} are the original row and column coordinates of the GCP in the image and x' and y' are the computed or estimated coordinates in the original image. The square root of the squared deviations represents a measure of the accuracy of this GCP in the image. By computing RMS_{error} for all GCPs, it is possible to 1) see which GCPs exhibit the greatest error, and 2) sum all the RMS_{error}.

Normally, the user specifies a certain amount (a threshold) of acceptable total RMS_{error} (e.g., 1 pixel). If an evaluation of the total RMS_{error} reveals that a given set of control points exceeds this threshold, it is common practice to 1) delete the GCP that has the greatest amount of individual error from the analysis, 2) recompute the six coefficients, and 3) recompute the RMS_{error} for all points. This process continues until one of the following occurs: the total RMS_{error} is less than the threshold specified or too few points remain to perform a least-squares regression to compute the coefficients. Once the acceptable RMS_{error} is reached, the analyst can proceed to the intensity interpolation phase of geometric rectification, which attempts to fill an output grid (*x, y*) with brightness values found in the original input grid (x', y').

Intensity Interpolation

The intensity interpolation process involves the extraction of a brightness value from an x', y' location in the original (distorted) input image and its relocation to the appropriate *x, y* coordinate location in the rectified output image. This pixel-filling logic is used to produce the output image line by line, column by column. Most of the time the x' and y' coordinates to be sampled in the input image are floating point numbers (i.e., they are not integers). For example, in Figure 7-8b we see that pixel 5, 4 (*x, y*) in the output image is to be filled with the value from coordinates 2.4, 2.7 (x', y') in the original input image. When this occurs, there are several methods of brightness value (*BV*) interpolation that can be applied, including:

- nearest neighbor,
- bilinear interpolation, and
- cubic convolution.

The practice is commonly referred to as *resampling*.

Nearest-neighbor Interpolation: When an image analyst uses *nearest-neighbor* interpolation, the brightness value closest to the specified x', y' coordinate is assigned to the output *x, y* coordinate. For example, in Figure 7-8b, the output pixel 5, 4 (*x, y*) requests the brightness value in the original input image at location 2.4, 2.7 (x', y'). There is no value at this location. However, there are nearby values at the integer grid intersections. Distances from 2.4, 2.7 (x', y') to neighboring pixels are computed using the Pythagorean

TABLE 7–2 Bilinear interpolation of a weighted brightness value (BV_{wt}) at location x', y' based on the analysis of four sample points in Figure 7-8b.

Sample Point Location (column, row)	Value at Sample Point, Z	Distance from x', y' to the Sample Point, D	D_k^2	$\frac{Z}{D_k^2}$	$\frac{1}{D_k^2}$
2, 2	9	$D = \sqrt{(2.4-2)^2 + (2.7-2)^2} = 0.806$	0.65	13.85	1.539
3, 2	6	$D = \sqrt{(2.4-3)^2 + (2.7-2)^2} = 0.921$	0.85	7.06	1.176
2, 3	15	$D = \sqrt{(2.4-2)^2 + (2.7-3)^2} = 0.500$	0.25	60.00	4.000
3, 3	18	$D = \sqrt{(2.4-3)^2 + (2.7-3)^2} = 0.670$	0.45	40.00	2.222
				Σ 120.91	Σ 8.937
				BV_{wt} = 120.91/8.937 =**13.53**	

theorem. A nearest-neighbor rule would assign the output pixel (x, y) the value of 15, which is the value found at the nearest input pixel.

This is a computationally efficient procedure. It is especially liked by Earth scientists because it does not alter the pixel brightness values during resampling. It is often the very subtle changes in brightness values that make all the difference when distinguishing one type of vegetation from another, an edge associated with a geologic lineament, or different levels of turbidity, chlorophyll, or temperature in a lake. Other interpolation techniques to be discussed use averages to compute the output intensity value, often removing valuable spectral information. Nearest-neighbor resampling should be used whenever biophysical information is to be extracted from remote sensor data.

Bilinear Interpolation: First-order or *bilinear interpolation* assigns output pixel values by interpolating brightness values in two orthogonal directions in the input image. It basically fits a plane to the four pixel values nearest to the desired position (x', y') in the input image and then computes a new brightness value based on the weighted distances to these points. For example, the distances from the requested x', y' position at 2.4, 2.7 in the input image in Figure 7-8b to the closest four input pixel coordinates (2, 2; 3, 2; 2, 3; 3, 3) are computed in Table 7-2. The closer a pixel is to the desired x', y' location, the more weight it will have in the final computation of the average. The weighted average of the new brightness value (BV_{wt}) is computed using the equation:

$$\text{Bilinear}_{BV_{wt}} = \frac{\sum_{k=1}^{4} \frac{Z_k}{D_k^2}}{\sum_{k=1}^{4} \frac{1}{D_k^2}}, \tag{7.14}$$

where Z_k are the surrounding four data point values, and D_k^2 are the distances squared from the point in question (x', y') to these data points. In our example, the weighted average of BV_{wt} is 13.53 (truncated to 13), as shown in Table 7-2. The average without weighting is 12. In many respects this method acts as a spatial moving filter that subdues extremes in brightness value throughout the output image.

Cubic Convolution: Resampling assigns values to output pixels in much the same manner as bilinear interpolation, except that the weighted values of sixteen input pixels surrounding the location of the desired x', y' pixel are used to determine the value of the output pixel. For example, the distances from the requested x', y' position at 2.4, 2.7 in the input image in Figure 7-8b to the closest 16 input pixel coordinates are computed in Table 7-3. The weighted average of the new brightness value (BV_{wt}) is computed using the equation:

$$\text{Cubic Convolution}_{BV_{wt}} = \frac{\sum_{k=1}^{16} \frac{Z_k}{D_k^2}}{\sum_{k=1}^{16} \frac{1}{D_k^2}}, \tag{7.15}$$

where Z_k are the surrounding 16 data point values, and D_k^2 are the distances squared from the point in question (x', y') to these data points. In this example, the weighted average of BV_{wt} is 13.41 (truncated to 13), as shown in Table 7-3. The average without weighting after truncation is 12.

TABLE 7–3 Cubic convolution interpolation of a weighted brightness value (BV_{wt}) at location x', y' based on the analysis of 16 sample points in Figure 7-8b.

Sample Point Location (column, row)	Value at Sample Point, *Z*	Distance from x', y' to the Sample Point, *D*	D_k^2	$\frac{Z}{D_k^2}$	$\frac{1}{D_k^2}$
1, 1	7	$D = \sqrt{(2.4-1)^2 + (2.7-1)^2} = 2.202$	4.85	1.443	0.206
2, 1	8	$D = \sqrt{(2.4-2)^2 + (2.7-1)^2} = 1.746$	3.05	2.623	0.328
3, 1	6	$D = \sqrt{(2.4-3)^2 + (2.7-1)^2} = 1.80$	3.24	1.852	0.309
4, 1	7	$D = \sqrt{(2.4-4)^2 + (2.7-1)^2} = 2.335$	5.45	1.284	0.183
1, 2	8	$D = \sqrt{(2.4-1)^2 + (2.7-2)^2} = 1.565$	2.45	3.265	0.408
2, 2	9	$D = \sqrt{(2.4-2)^2 + (2.7-2)^2} = 0.806$	0.65	13.85	1.539
3, 2	6	$D = \sqrt{(2.4-3)^2 + (2.7-2)^2} = 0.921$	0.85	7.06	1.176
4, 2	14	$D = \sqrt{(2.4-4)^2 + (2.7-2)^2} = 1.746$	3.05	4.59	0.328
1, 3	14	$D = \sqrt{(2.4-1)^2 + (2.7-3)^2} = 1.432$	2.05	6.829	0.488
2, 3	15	$D = \sqrt{(2.4-2)^2 + (2.7-3)^2} = 0.500$	0.25	60.00	4.000
3, 3	18	$D = \sqrt{(2.4-3)^2 + (2.7-3)^2} = 0.670$	0.45	40.00	2.222
4, 3	17	$D = \sqrt{(2.4-4)^2 + (2.7-3)^2} = 1.63$	2.65	6.415	0.377
1, 4	14	$D = \sqrt{(2.4-1)^2 + (2.7-4)^2} = 1.911$	3.65	3.836	0.274
2, 4	16	$D = \sqrt{(2.4-2)^2 + (2.7-4)^2} = 1.360$	1.85	8.649	0.541
3, 4	19	$D = \sqrt{(2.4-3)^2 + (2.7-4)^2} = 1.432$	2.05	9.268	0.488
4, 4	20	$D = \sqrt{(2.4-4)^2 + (2.7-4)^2} = 2.062$	4.25	4.706	0.235
				Σ 175.67	Σ 13.102

$BV_{wt} = 175.67 / 13.102$
$BV_{wt} = \mathbf{13.41}$



An Example of Image-to-Map Rectification

To appreciate digital *image-to-map* rectification, it is useful to demonstrate the logic by applying it to a real dataset such as a Landsat Thematic Mapper image of Charleston, SC. The image-to-map rectification process generally involves:

- selecting an appropriate map projection,
- collecting ground control points,

- determining the optimum set of geometric rectification coefficients by iteratively computing the total GCP RMS_{error}, and
- filling the output matrix using spatial and intensity interpolation resampling methods.

Selecting an Appropriate Map Projection

A **map projection** is a systematic transformation of the three-dimensional Earth into a two-dimensional flat map (Iliffe, 2008; Garnett, 2009). The characteristics of the map projection to which the remote sensor data will be rectified are a very important consideration (Merchant and Narumalani, 2009). Cartographers and mathematicians have devised a great variety of map projections to project (flatten) the three-dimensional characteristics of the globe onto a two-dimensional flat map (Robinson and Snyder, 1991; Maher, 2010). Many of the most useful map projections are summarized in John Snyder's USGS Professional Paper *Map Projection: A Working Manual* (1987) and in *Flattening the Earth* (1995). The following material identifies some of the key properties, characteristics, and preferred uses of several of the most important map projections frequently used by image analysts for image rectification.

Developable Surfaces used to Create Map Projections

Cartographers have developed methods to project the information found on the surface of a globe onto simple geometric forms, called developable surfaces, including (Figure 7-10):

- planes,
- cylinders, and
- cones.

A **developable surface** is a simple geometric form capable of being flattened without compressing or stretching (Slocum et al., 2008). A plane is already flat, whereas a cylinder or cone may be cut and laid out flat, without stretching or compressing the content. Map projections are typically classified into three general families: cylindrical, conical, and azimuthal (planar).

One of the most important characteristics of map projections is where the developable surface just touches or intersects the sphere. For example, if the sphere (globe) just touches the developable surface at a single point, it is said to be *tangent* to the surface as shown in Figure 7-10a,c,e. If the sphere intersects the developable surface it is said to be *secant* to the surface as shown in Figure 7-10b,d,f). The place(s) where the sphere intersects the developable surface (i.e., it is either tangent or secant to the surface) is the most accurate area of a map projection.

Map Projection Characteristics

Unfortunately, the map projection process cannot be accomplished without some distortion, resulting in every map projection having unique advantages and disadvantages (Grafarend and Krumm, 2006; Kanters, 2007; Krygier, 2011) (Table 7-4). Consequently, there is no universal "best" map projection for each mapping project (USGS, 2011). Instead, the image analyst must select the map projection best suited to his or her needs, which minimizes the distortion of the most important features to be portrayed on the map.

Please remember that every two-dimensional flat map distorts the surface of the three-dimensional Earth in some way. A projected flat map or parts of a map can have some, but never all, of the following characteristics (Table 7-4):

- true directions,
- true distances (i.e., equidistant),
- true areas, and/or
- true shapes (i.e., conformality).

A map projection is said to be **conformal** when at any point on the map the scale is the same in every direction (USGS, 2011). Therefore, meridians and parallels in conformal map projections intersect at right angles and the shapes of very small areas and angles with very short sides are preserved. The size of most areas, however, is distorted. Because angles at each point are correct on conformal maps, the scale in every direction around any point is constant. This allows the analyst to measure distance and direction between relatively near points with good precision. For our purposes, this means that for image areas covering a few contiguous 1:24,000-scale 7.5-minute quadrangle sheets, accurate spatial measurement is possible if the data are rectified to a conformal map projection.

A map projection is **equal-area** (also referred to as equivalent) if every part on the map, as well as the whole, has the same area as the corresponding part on the Earth, at the same reduced scale. No flat map can be both equal-area and conformal at the same time (USGS, 2011).

Equidistant maps show true distances only from the center of the projection or along a special set of lines (USGS, 2011). They correctly represent azimuths about the point of tangency (i.e., where a map touches the globe) but do not correctly represent areas. For example, an Azimuthal Equidistant map projection centered on Washington, DC, shows the correct distance between Washington, DC, and any other point in the

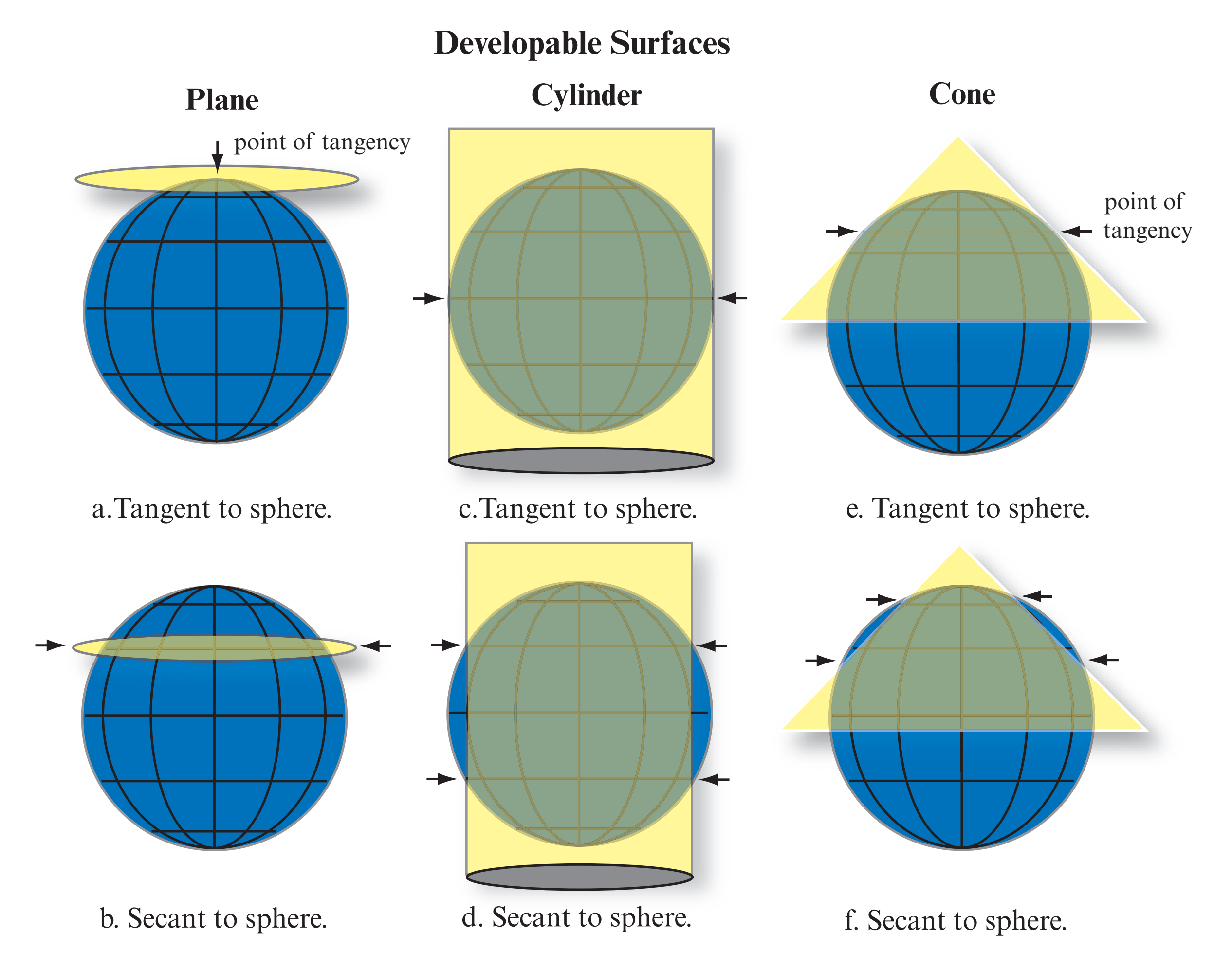


FIGURE 7–10 Three types of developable surfaces are often used to create map projections: plane, cylinder, and cone. ab) A planar developable surface may just touch (i.e., be tangent with) or intersect (i.e., be secant with it) a sphere. cd) A cylindrical developable surface may be tangent or secant to a sphere. ef) A cone developable surface may be tangent or secant to a sphere (Jensen and Jensen, 2013).

projection, i.e., it shows the correct distance between Washington, DC, and Philadelphia, PA, and between Washington, DC, and Richmond, VA. But it does not show the correct distance between Philadelphia, PA, and Richmond, VA, because the projection is not centered on Philadelphia, PA. No flat map can be both equidistant and equal-area at the same time. A circle of diameter *n* drawn at any location on the map will encompass exactly the same geographic area. This characteristic is useful if a scientist is interested in comparing land-use area, density, and so on. Unfortunately, to maintain the equal-area attribute, the shapes, angles, and scale in parts of the map may be distorted.

Azimuthal map projections correctly represent selected angular relationships. **Azimuthal** projections are projections to a plane placed *tangent* to (just touching) the globe at a point. As with distances, not all angular relationships can be represented correctly on a single map, but it is possible to correctly represent all angular relationships about a single point. A basic knowledge of the properties of commonly used map projections helps when selecting a map projection that comes closest to fulfilling a specific need.

It is not possible to describe or show examples of all the map projections available (Esri, 2004; Intergraph, 2013; Furuti, 2011). Rather, the following discussion focuses on just a few map projections that are especially useful when rectifying remote sensor data.

Cylindrical Map Projections

Standard cylindrical map projections are mathematically projected onto a cylinder that a) just touches the sphere (i.e., it is tangent to the sphere), or b) is made to intersect the sphere (i.e., it is secant to the sphere) (Figure 7-11ab).

Mercator Map Projection: The Mercator map projection was developed by Flemish cartographer Gerardus Mercator (1512–1594) in 1569 for navigation purposes (Figure 7-12ab). Meridians and parallels are

TABLE 7–4 Summary of selected map projection properties associated with cylindrical, azimuthal, and conic developable surfaces. P = partly (based on USGS, 2011).

Developable Surface Type	Projection	Conformal	Equal area	Equi-distant	True Direction	Perspective	Straight Rhumb lines
Sphere	Globe						
Cylindrical	Mercator				P		
	Transverse Mercator						
	Space Oblique Mercator						
Azimuthal	Gnomonic				P		
	Stereographic				P		
	Orthographic				P		
	Azimuthal Equidistant			P	P		
	Lambert Azimuthal Equal-area				P		
Conic	Albers Equal-area						
	Lambert Conformal				P		
	Equidistant			P			
	Polyconic			P			

straight lines and cross one another at 90°. Angular relationships are maintained in this map projection. However, to preserve conformality, parallels of latitude are placed increasingly farther apart with increasing distance from the Equator (Figure 7-12ab). Any straight line on a Mercator map projection is a **rhumb line** (a line of constant angular direction), which is not necessarily the shortest distance between points. Distances are true only along the Equator (i.e., the standard parallel), but are reasonably correct within 12° to 15° north or south of the Equator (Figure 7-11a). If the developable surface is secant to the sphere, two standard parallels will have the correct scale instead of the Equator (Figure 7-11b) (Slocum et al., 2008; Krygier, 2011).

Areas and shapes of large areas like continents are usually distorted in Mercator map projections. The distortion increases away from the Equator and there is extreme distortion in the polar regions. The poles are typically not shown in a Mercator map projection because of the extreme distortion in these areas. A Mercator map projection of MODIS satellite image data and global elevation data are shown in Figure 7-12ab.

The Mercator map projection is excellent for mapping equatorial regions. Otherwise, the Mercator projection should be considered a special-purpose map projection best suited for navigation. Secant Mercator map projections are used for large-scale coastal charts. In fact, the use of the Mercator map projection for nautical charting is universal. Examples are the charts published by the National Ocean Survey, U.S. Dept. of Commerce (Intergraph, 2013).

Mercator map projections are conformal in that angles and shapes within any small geographic area (such as that found in a typical USGS 7.5-minute topographic

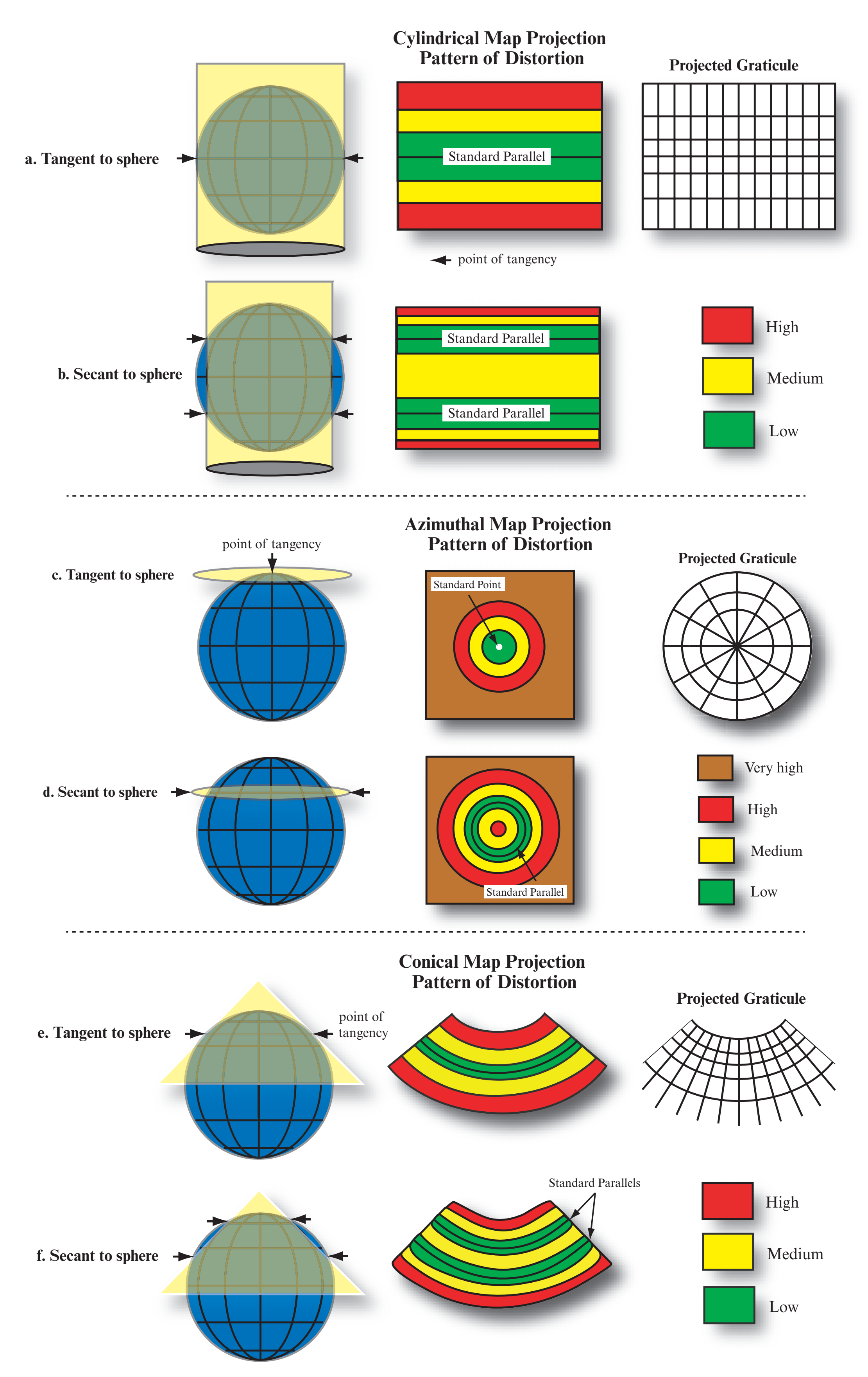
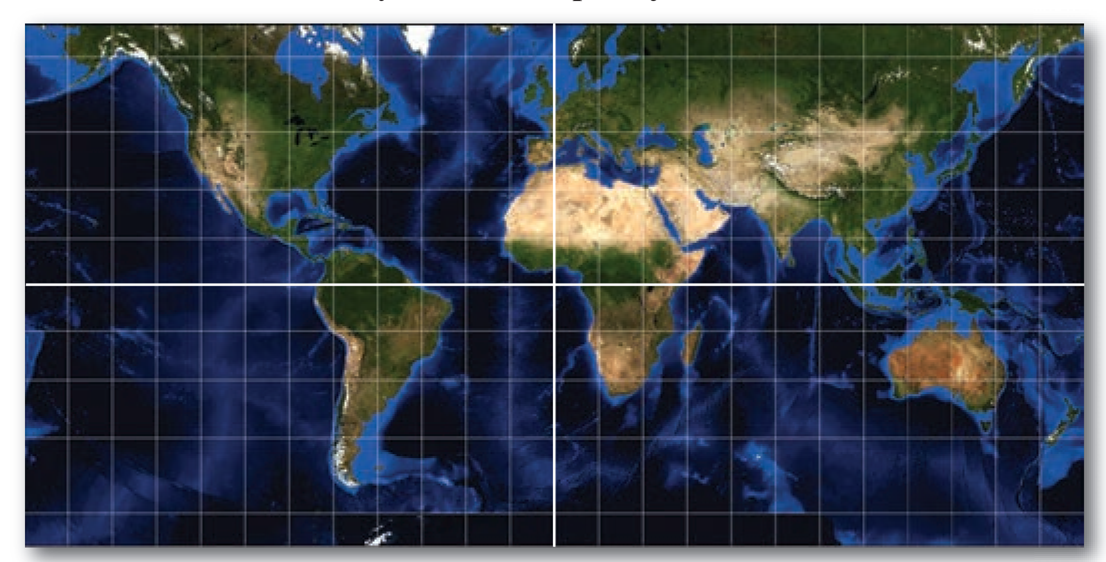
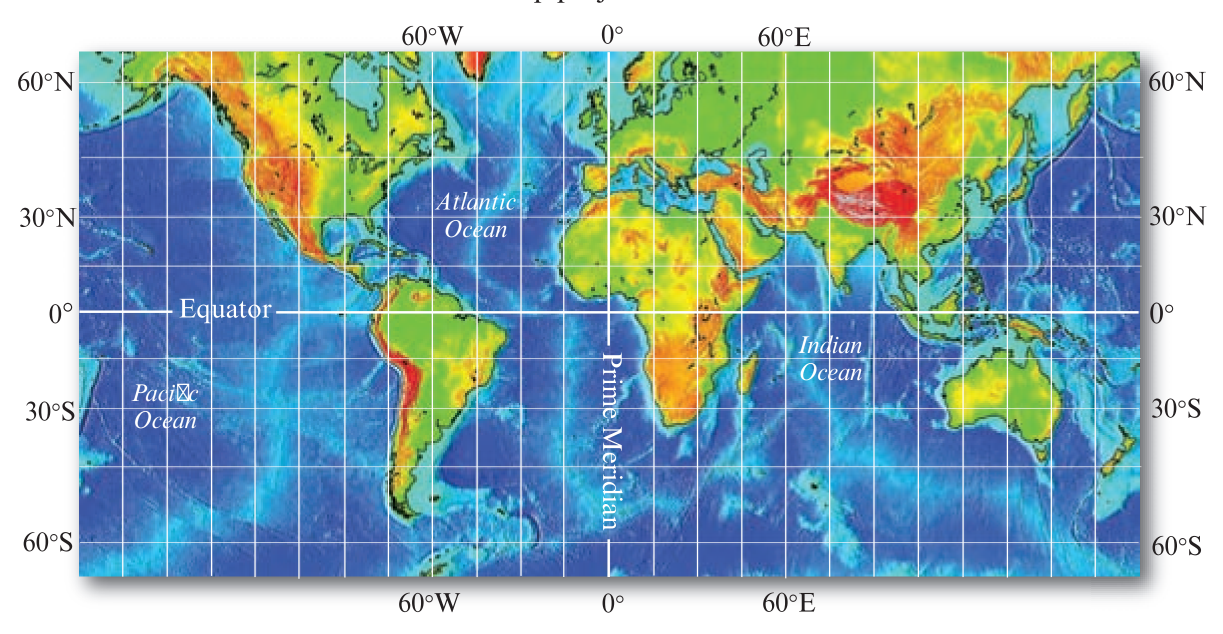


FIGURE 7–11 Cylindrical map projection pattern of distortion where a) the developable surface just touches the sphere, i.e., it is tangent, and b) when it intersects the sphere, i.e. it is secant. cd) Tangent and secant azimuthal map projection patterns of distortion. ef) Tangent and secant conical map projection patterns of distortion (adapted from Jensen and Jensen, 2013).

Cylindrical Map Projections



a. Mercator conformal map projection of MODIS satellite data.



b. Mercator conformal map projection of elevation data.

FIGURE 7–12 a) A Mercator map projection of NASA Moderate Resolution Imaging Spectrometer (MODIS) satellite data. The Mercator conformal map projection is based on a cylindrical developable surface (refer to Figure 7-11ab). The Equator is often the standard parallel. Note the extreme distortion of land masses at higher latitudes. b) A Mercator conformal map projection of global elevation data. The MODIS imagery and elevation data are courtesy of NASA. Both projections were created using NASA's *G.Projector* software.

map) are essentially true (USGS, 2011). The patterns of distortion associated with cylindrical map projections that just touch (tangent) or intersect (secant) the globe are shown in Figure 7-11ab (Slocum et al., 2008; Krygier, 2011).

Universal Transverse Mercator (UTM) Projection: One of the most widely used map projections for rectifying remotely sensed data and for large-scale topographic mapping applications is the Transverse Mercator projection. It is made from a normal Mercator projection by rotating the cylinder (the developable surface) 90° so that it lies tangent along a *meridian* (line of longitude) instead of line of latitude.

The Universal Transverse Mercator (UTM) coordinate system is based on a cartesian coordinate system. It is composed of 60 zones, each 6° of longitude wide, with

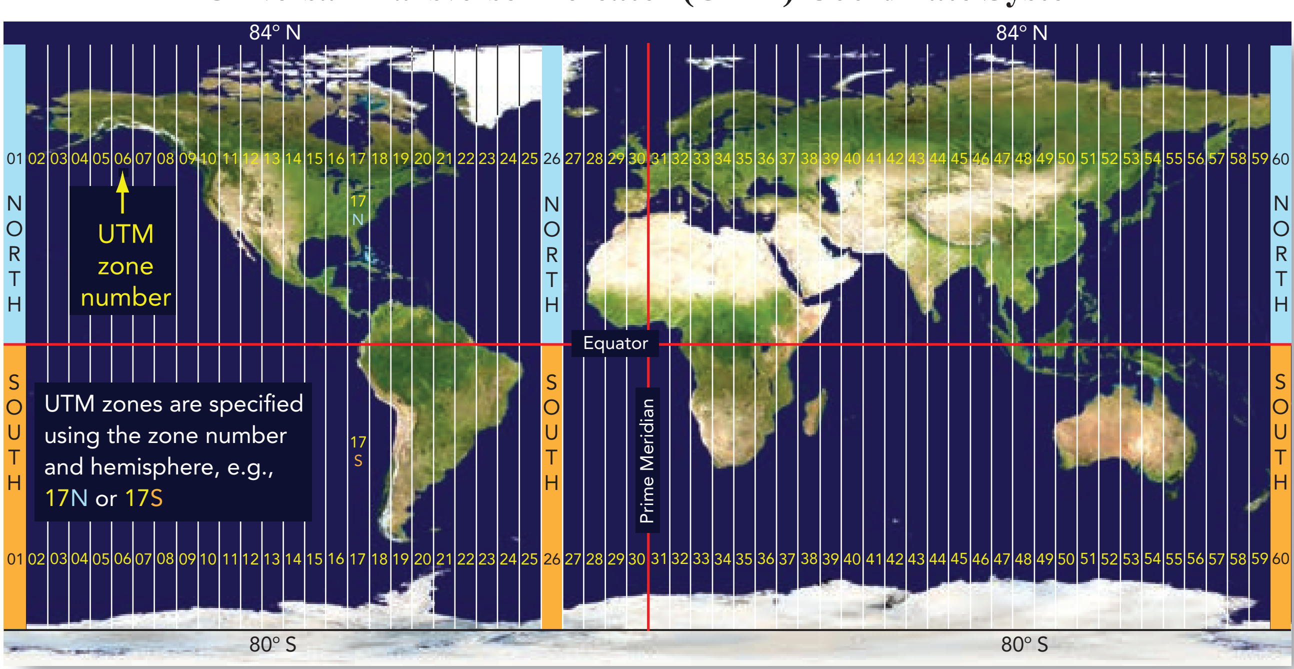


FIGURE 7–13 There are 60 UTM zones, each 6° of longitude wide from zone 01 at 180° W of the Prime Meridian to zone 60 at 180° E. Each UTM zone has both a north and a south zone. The north zone extends from the Equator to 84° N latitude. The south zone extends from 80° S latitude to the Equator (NGA, 2007) (MODIS imagery courtesy of NASA Goddard Space Flight Center; Reto Stöckli).

a Central Meridian placed every sixth meridian beginning with 177° West longitude. Zone 1 is located from 180° to 174° West longitude. Zone 2 extends from 174° west to 168° West longitude and so forth. Each UTM zone has both a north and a south zone (Figure 7-13a). The north zone extends from the Equator to 84° N latitude. The southern zone extends from 84° S latitude to the Equator (Figure 7-13 and 7-14a). The UTM map projection is not used to map the poles.

The intersection of the Central Meridian of a UTM and the Equator is the origin of the zone (Figure 7-14a). It has a central scale factor of 0.9996. The Central Meridians, the Equator, and each line 90° from the Central Meridian are straight lines (Figure 7-14a). The Central Meridian normally has a constant scale. Any lines parallel to the Central Meridian are lines of constant scale. UTM maps can be edge-matched only if they are in the same zone with one Central Meridian.

Each UTM zone has a Central Meridian that remains perpendicular to the zone's southern boundary throughout the north-south extent of the zone. Each zone also has a false-easting of 500,000 m west of the zone's Central Meridian. This ensures that all easting and northing values remain positive throughout the zone. Otherwise, you would have negative values west of the Central Meridian.

The UTM projection is often used for mapping large areas that are mainly north-south in extent. For example, the UTM projection is often used by the U.S. Geological Survey in its topographic mapping program at scales from 1:24,000 to 1:250,000. UTM Zone 17 N is useful for mapping South Carolina with a Central Meridian of 81° W (Figure 7-14b). UTM Zone 16 N is useful for mapping Alabama and is centered on 87° W longitude (Figure 7-14b).

Distances are true only along the Central Meridian selected by the image analyst or along two lines parallel to it, but all distances, directions, shapes, and areas are reasonably accurate within 15° of the Central Meridian. Distortion of distances, directions, and size of areas increase rapidly outside the 15° band. Because the map is conformal, however, shapes and angles within any small area (such as that found on a 1:24,000-scale USGS topographic map) are essentially true.

UTM coordinates are in meters and referenced according to the x (easting) and y (northing) values. Easting refers to the meters east of the artificial zone boundary. Northing refers to meters north of the artificial zone boundary. In the northern hemisphere, northing refers to meters north of the Equator. For example, suppose that you are given the following UTM coordinate information:

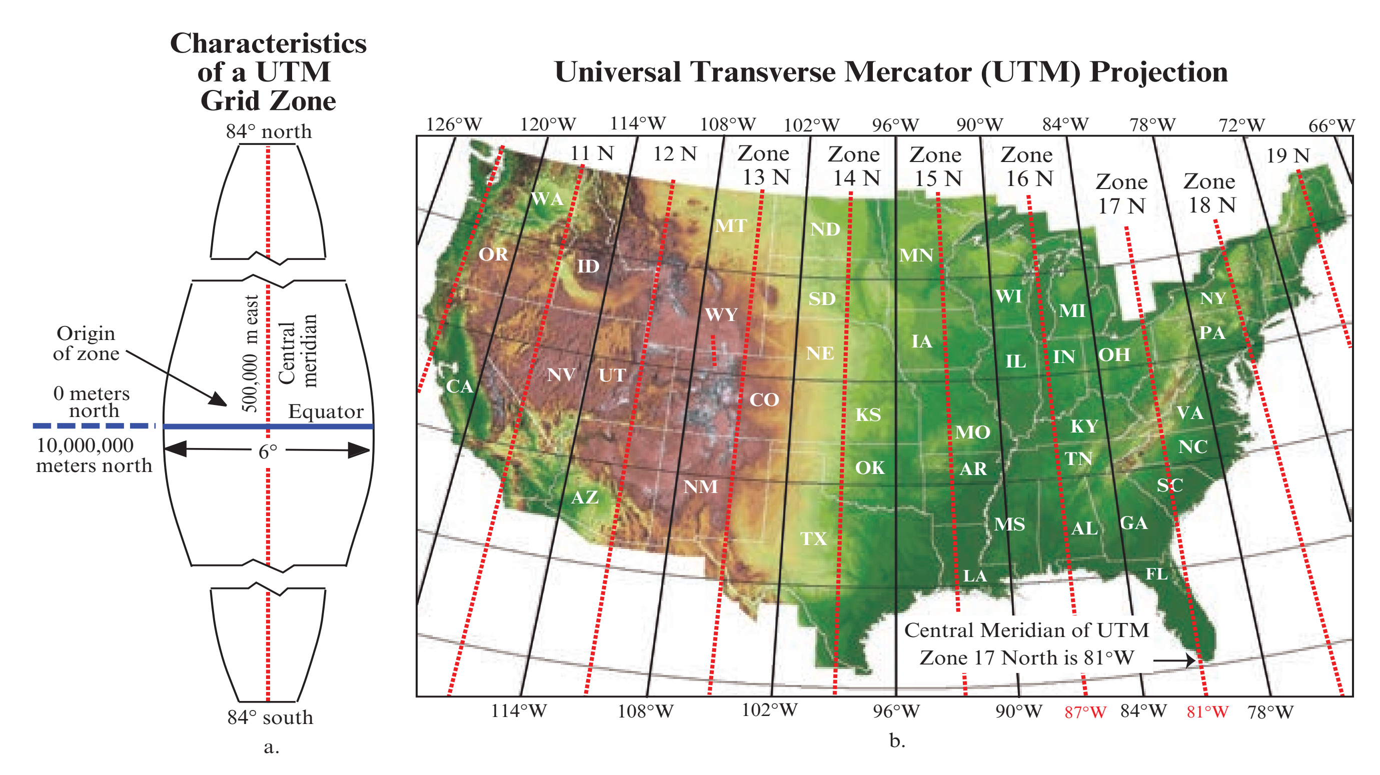


FIGURE 7–14 Universal Transverse Mercator (UTM) grid zone with associated parameters. This projection is often used when rectifying remote sensor data to a base map. It is found on U.S. Geological Survey 7.5- and 15-minute quadrangles. a) Characteristics of a UTM Grid Zone. There are 60 zones, each 6° in size encompassing the Earth. b) UTM grid Zone 17 N with a Central Meridian at 81°W longitude is useful for geospatial projects in South Carolina. UTM Zone 16 N centered on 87°W longitude is useful for projects in Alabama. Elevation data courtesy of the U.S. Geological Survey.

12N 444782 E 4455672 N

This means that the location is in Zone 12 north of the Equator, 444,782 meters east of the UTM origin and 4,455,672 meters north of the Equator. Of course, you can also convert this location to other units. For example, you can divide these values by 1,000 to convert the units to kilometers. In this case, you could determine that the location is 444.782 km east of the UTM origin and 4,455.672 km north of the Equator.

UTM coordinates are easy to use and virtually all GPS devices allow for locations to be mapped in UTM coordinates. In addition, the UTM coordinate system covers most of the world (except for the poles). Also, because the system uses meters as its base unit, it is easy to convert to other units and to calculate useful areal units such as meters^2, hectares, and kilometers^2. Internet mapping applications, such as *Google Earth*, allow for coordinates to be displayed and queried in UTM.

Some disadvantages to the UTM coordinate system must be noted. It is often difficult to use a UTM map projection when the study area crosses multiple UTM zones from east to west. Therefore, if your study area is located across two UTM zones, a UTM map projection may not be the best choice. This makes using UTM very difficult for large, continental-sized areas such as all of the conterminous United States or the entire area of the Amazon Basin.

Space Oblique Mercator Projection: The useful Space Oblique Mercator (SOM) projection is a modified cylindrical projection with the map surface defined by a satellite orbit. It was developed by U.S. Geological Survey scientists in the 1970s to reduce the amount of distortion caused when satellite images of the ellipsoidal Earth were printed on a flat page. It was originally developed to map the imagery obtained by the Landsat Multispectral Scanner (MSS). The projection is used to map the continuous swath of remote sensing data collected by a satellite during each orbital pass. The scale is true along the ground track (Snyder, 1987; 1995).

The SOM projection is useful mainly for the relatively narrow swath along the groundtrack of the satellite. SOM maps are basically conformal with the extent of the map defined by the orbital characteristics of the remote sensing system. The SOM may be used for any satellite orbiting the Earth in a circular or elliptical orbit and at any inclination.

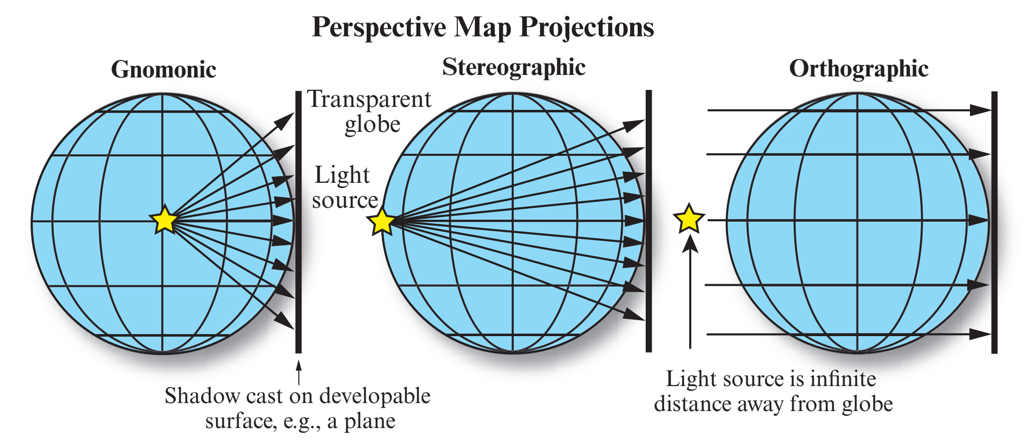


FIGURE 7–15 Perspective map projections: Gnomonic, Stereographic, and Orthographic. Note the location of the light source in relation to the location of the transparent globe and the developable surface, which in this particular example is a plane. Also note the angle of the shadows cast on the developable surface (Jensen and Jensen, 2013).

Azimuthal (Planar) Map Projections

Azimuthal map projections are mathematically projected onto a plane *tangent* to (touching) any point on the globe, such as at the North Pole (90°N, 0°W) (Figure 7-11c). They can also be created by intersecting the sphere with a plane, e.g., along 40° N (Figure 7-11d). The patterns of distortion associated with azimuthal map projections that just touch (tangent to) or intersect (secant to) the globe are shown in Figure 7-11cd (Slocum et al., 2008; Krygier, 2011). Azimuthal map projections may be centered on the poles (referred to as Polar-aspect), at a point on the Equator (referred to as Equatorial-aspect), or at any other orientation desired (referred to as Oblique-aspect).

Perspective Azimuthal Map Projections: Perspective Azimuthal map projections may be constructed geometrically using a light source placed at a certain location within or outside of a transparent globe that contains parallels of latitude and meridians of longitude (Figure 7-15). Rays of light from the light source project shadows of the meridians and parallels onto the planar surface.

The location of the light source can be inside the sphere (referred to as *gnomonic*), on the opposite side of the sphere (the antipode) at the point of tangency (referred to as *stereographic*), or at an infinite distance from the point of tangency, yielding parallel light rays (referred to as *orthographic*) (Figure 7-15). In reality, the graticule (grid) of the globe is projected mathematically onto the plane rather than by using light projected through the sphere.

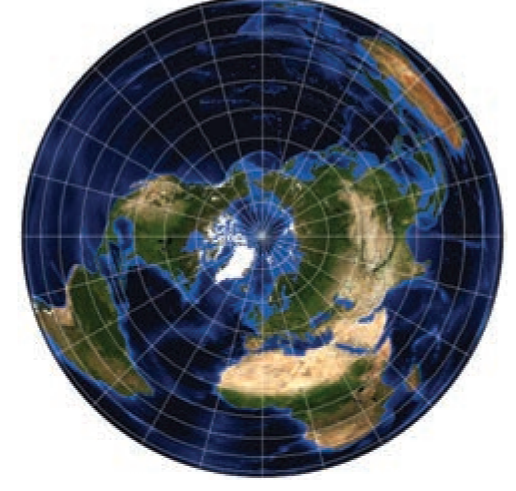
The Azimuthal Gnomonic projection is believed to be the oldest projection, developed around the sixth century B.C. Its most important property is that the Equator and all meridians are mapped as straight lines, making it possible to find the shortest route between any two points.

Azimuthal Stereographic maps are often used to display the hemisphere that is opposite the light source point. It is not possible to show both hemispheres in their entirety. It is the only azimuthal projection that preserves true angles and local shape. It is often used to map large continent-sized areas of similar extent in all directions (USGS, 2011).

The Azimuthal Orthographic projection is often used to display the Earth, moon, and other planets as if the viewer were infinitely far away in outer space. The display closely resembles a three-dimensional view of the spheres. It is the most familiar of the azimuthal map projections (Intergraph, 2013).

Lambert Azimuthal Equal-area: The Lambert Azimuthal Equal-area map projection (Figure 7-16) is best suited for regions extending equally in all directions from a standard point of tangency. It is possible to locate the point of tangency at the poles (e.g., Figure 7-16a) or anywhere else the user desires such as at 40°N, 0°W longitude (Figure 7-16b) or at 40°N, 80°W longitude as demonstrated using MODIS imagery of the Earth (Figure 7-16c).

Areas on the map are shown in their true proportion to the same areas on the Earth. Therefore, any two quadrangles that are bounded by two parallels and two meridians at the same latitude are equal in area. Directions are true only from the standard point of tangency. The scale decreases (becomes smaller) gradually away from the standard point of tangency. The distortion of shapes increases away from the standard point of tangency. Any straight line drawn through the standard point is on a great circle. The map is equal-



a. Lambert Azimuthal Equal-area 90° N, 0° W.



b. Lambert Azimuthal Equal-area 40° N, 0° W.



c. Lambert Azimuthal Equal-area 40° N, 80° W.

FIGURE 7–16 Selected Lambert Azimuthal Equal-area map projections. a) Lambert Azimuthal Equal-area map projection of MODIS satellite imagery with the standard point of tangency at 90° N, 0° W. b) Lambert Azimuthal Equal-area map projection of MODIS data with the point of tangency at 40° N, 0° W. c) The point of tangency is at 40° N, 80° W (based on the use of NASA map projection software and NASA MODIS imagery).

area but not conformal, perspective, or equidistant (USGS, 2011).

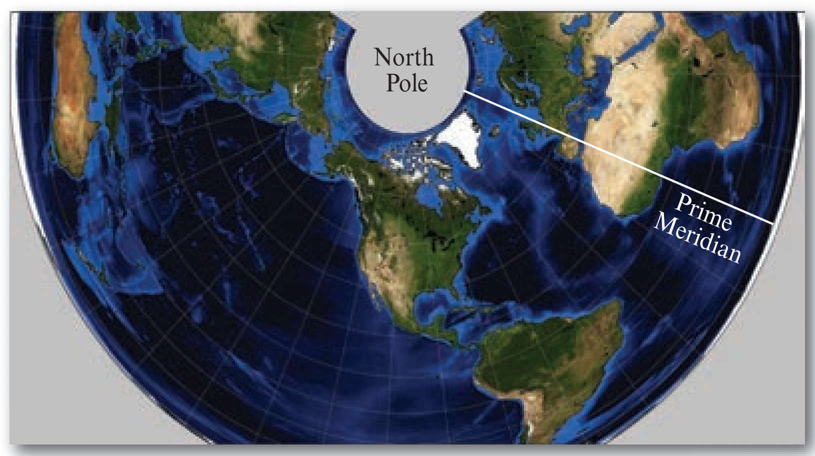
Azimuthal Equidistant: In an Azimuthal Equidistant map projection, distances and directions to all places are true only from the standard point of tangency of the projection. Distances are correct between points along straight lines through the standard (center) point. All other distances are inaccurate. Any straight line drawn through the standard (center) point is on a great circle. The distortion of areas and shapes increases away from the standard point (USGS, 2011). Azimuthal Equidistant projections are useful for showing airline distances from the standard point of the projection. A polar aspect is often used for world maps and maps of polar hemispheres. An oblique aspect is used for atlas maps of continents and world maps for aviation use.

Conical Map Projections

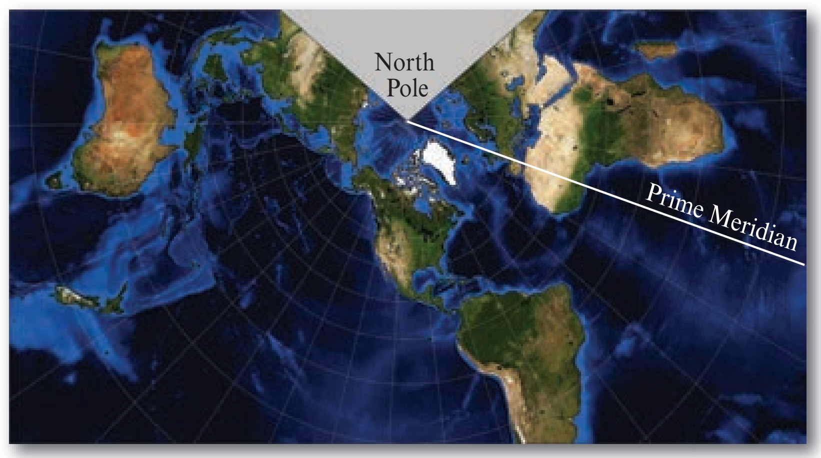
Conical map projections are mathematically projected onto a cone-developable surface. The cone can be tangent or secant to the sphere. If it is tangent, there is one standard parallel. If it is secant, there are two standard parallels (Figure 7-11ef). The greater the distance from the point of tangency or intersection of the sphere with the cone, the greater the distortion (Slocum et al., 2008; Krygier, 2011).

Albers Equal-area Conic: The Albers Equal-area Conic map projection is used by the U.S. Geological Survey for maps showing the conterminous United States (i.e., the 48 mainland states) or for mapping

Conical Map Projections



a. Albers Equal-area Conic.



b. Lambert Conformal Conic.

FIGURE 7–17 Selected conical map projections. a) Albers Equal-area Conic map projection. b) Lambert Conformal Conic map projection (based on the use of NASA map projection software and NASA MODIS imagery).

large areas of the United States (Figure 7-17a). The map projection is secant at two standard parallels. The projection is ideal when mapping large areas that are mainly east-west in geographic extent and when it is important to maintain equal-area relationships. The projection is not conformal, perspective, or equidistant. Directional measurement is reasonably accurate within limited geographic regions. Distance measurements are accurate along both of the standard parallels. The scale is true only along the two standard parallels. Maps created using the Albers Equal-area Conic projection can be edge-matched only if they have the same standard parallels and the same scale (USGS, 2011).

Lambert Conformal Conic: The Lambert Conformal Conic map projection (Figure 7-17b) is one of the most widely used map projections. It is secant at two standard parallels. It looks like the Albers Equal-area Conic (Figure 7-17a), but the graticule spacing is different. The projection is used by the USGS for many 7.5- and 15-minute topographic maps and for the State Base Map series. It is also used to map countries or regions that are mainly east-west in extent (USGS, 2011).

The Lambert Conformal Conic projection is not perspective, equal-area, or equidistant. Distance measurement is true only along the two standard parallels and reasonably accurate elsewhere in limited regions. Directional measurements are reasonably accurate, especially near the standard parallels. The distortion of shapes and areas is minimized along the two standard parallels selected by the map maker, but increases the greater the distance away from the standard parallels. The shapes of areas on large-scale maps of relatively small geographic areas are essentially true.

The two standard parallels are 33°N and 45°N for the USGS Base Map series for the 48 conterminous States. The standard parallels are varied for the very important USGS Topographic Map series (both 7.5- and 15-minute) (USGS, 2011).

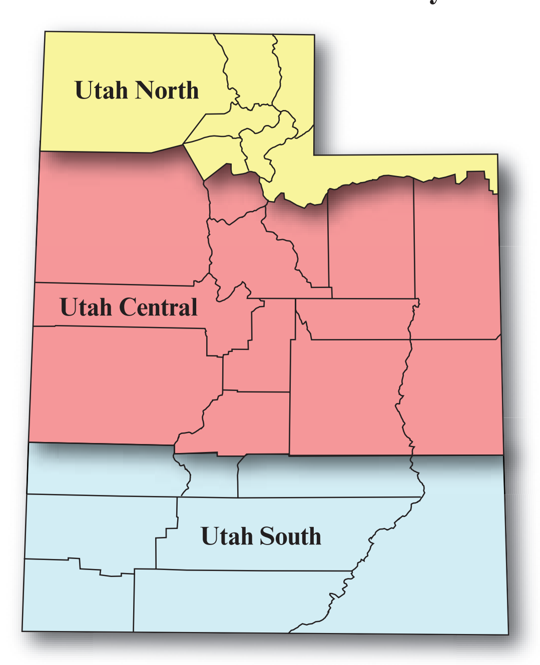


FIGURE 7–18 Each state has its own State Plane Coordinate System. The size and shape of each state determines the number of zones and their orientation. Utah has three parallel State Plane Coordinate zones (North, Central, and South).

Selected map projections ideally suited to mapping different geographic areas are summarized in Table 7-4. The globe is best at portraying the world. Azimuthal projections are especially useful for mapping entire hemispheres, continents, and regions. The cylindrical Transverse Mercator and several conic map projections are especially useful for medium- and large-scale mapping projects.

Other Projections and Coordinate Systems Useful for Image Rectification

Sometimes map projections or coordinate systems are specially prepared for political or other considerations. For example, in the 1930s each of the states agreed to develop their own map projection and coordinate system based on either the Transverse Mercator or the Lambert Conformal Conic map projections. These projections and coordinate systems came to be known as the State Plane Coordinate Systems (SPCS).

State Plane Coordinate Systems: In the U.S., individual state projections were chosen to decrease distortion and minimize error. These projections were chosen based on an individual state's shape and geographic location on the sphere. In addition, some states decided to break up their State Plane Coordinate System into different zones. For example, Utah's State Plane Coordinate System contains three different zones (North, Central, and South) based on the state's geographic shape and location (Figure 7-18). Utah's State Plane Coordinate System uses a Lambert Conformal Conic projection.

Most local government thematic data layers are based on the use of State Plane Coordinate Systems. Most legal descriptions used by governments include details in State Plane Coordinate System coordinates. So, if you plan on working for a local city or government agency you should be prepared to use the State Plane Coordinate System.

As with UTM, there are disadvantages to working with the State Plane Coordinate System. Similar to UTM, the State Plane Coordinate System used in most states has multiple zones that are difficult to move across if your study area happens to straddle the zones. In addition, State Plane Coordinate Systems end at the state line. Therefore, any projects that wish to examine multiple states will need to select another coordinate system. The original unit of measurement for the State Plane Coordinate System was based on the NAD27 datum in feet. Recently, the U.S. Geological Survey began releasing metric topographic maps. These maps are based on a new State Plane Coordinate System that is based on the NAD83 datum, with meters as the measurement unit.

Ground Control Point Collection

Twenty ground control points were located on the map and the UTM easting and northing of each point were identified (Table 7-5). The same 20 GCPs were then identified in the TM data according to their row and column coordinates (Table 7-5). The location of points 13, 14, and 16 is shown in Figure 7-5. The GCPs should be located uniformly throughout the region to be rectified and not congested into a small area simply because a) there are more easily identifiable points in that area or b) the locations are easy to get to in the field.

Determine Optimum Geometric Rectification Coefficients by Evaluating GCP Total RMS_{error}

The 20 GCPs selected were input to the least-squares regression procedure previously discussed to identify 1) the coefficients of the coordinate transformation, and 2) the individual and total RMS_{error} associated with the GCP s. A threshold of 0.5 was not satisfied until 13 GCPs were deleted from the analysis. The order in which the 13 GCPs were deleted and the RMS_{error} total found after each deletion are summarized in Table 7-5. The seven GCPs finally selected that produced an acceptable RMS_{error} are shown in Table 7-6. GCP 11 would have been the next point deleted if the threshold

TABLE 7–5 Characteristics of 20 ground control points used to rectify the Charleston, SC, Landsat Thematic Mapper scene.

Point Number	Order of Points Deleted[a]	Easting on Map, X_1	Northing on Map, X_2	X' Pixel	Y' Pixel	Total RMS_{error} after This Point Is Deleted
1	12	597,120	3,627,050	150	185	0.501
2	9	597,680	3,627,800	166	165	0.663
3	Kept	598,285	3,627,280	191	180	—
4	Kept	595,650	3,627,730	98	179	—
5	2	596,750	3,625,600	123	252	6.569
6	13	597,830	3,624,820	192	294	0.435
7	Kept	596,250	3,624,380	137	293	—
8	Kept	602,200	3,628,530	318	115	—
9	Kept	600,350	3,629,730	248	83	—
10	5	600,680	3,629,340	259	93	1.291
11	Kept	600,440	3,628,860	255	113	—
12	10	599,150	3,626,990	221	186	0.601
13	8	600,300	3,626,030	266	211	0.742
14	6	598,840	3,626,460	211	205	1.113
15	3	598,940	3,623,430	214	295	4.773
16	Kept	600,540	3,626,450	272	196	—
17	4	596,985	3,629,350	134	123	1.950
18	7	596,035	3,627,880	109	174	0.881
19	11	600,995	3,630,000	269	71	0.566
20	1	601,700	3,632,580	283	12	8.542
				Total RMS_{error} with all 20 GCPs used:		**11.016**

a. For example, GCP 20 was the first point deleted. After it was deleted, the total RMS_{error} dropped from 11.016 to 8.542. Point 5 was the second point deleted. After it was deleted, the total RMS_{error} dropped from 8.542 to 6.569.

had not been satisfied. The six coefficients derived from the seven suitable GCPs are found in Table 7-7. The Charleston Landsat TM scene rectified using these parameters is shown in Figure 7-6a.

It is instructive to demonstrate how the six coefficients were computed for this Landsat TM dataset. This was accomplished using only the final seven GCPs, as these represent the set that produced an acceptable RMS_{error} of <0.50 pixels. Remember, however, that this same operation was performed with 20 points, then 19 points, and so on, before arriving at just 7 acceptable points.

A least-squares multiple regression approach is used to compute the coefficients. Two equations are necessary. One equation computes the image y' coordinate (the dependent variable) as a function of the x and y map coordinates (representing two independent variables). The second computes the image x' coordinate as a function of the same map (x, y) coordinates. Three coefficients are determined by using each algorithm. The mathematics for computing the column coordinates (x') in the image can now be presented.

Multiple Regression Coefficients Computation

First, we will let

Y = either the x' or y' location in the image, depending on which is being evaluated; in this example it will represent the x' values

X_1 = easting coordinate (x) of the map GCP

TABLE 7–6 Information concerning the final seven ground control points used to rectify the Charleston, SC, Landsat Thematic Mapper Scene.

Point Number	Easting on Map	Adjusted Easting, X_1[a]	Northing on Map	Adjusted Northing, X_2[b]	Y[c]	
					X' Pixel	Y' Pixel
3	598,285	2,635	3,627,280	2,900	191	180
4	595,650	0	3,627,730	3,350	98	179
7	596,250	600	3,624,380	0	137	293
8	602,200	6,550	3,628,530	4,150	318	115
9	600,350	4,700	3,629,730	5,350	248	83
11	600,440	4,790	3,628,860	4,480	255	113
16	600,540	4,890	3,626,450	2,070	272	196
	Minimum = 595,650	24,165	Minimum = 3,624,380	22,300	1,519	1,159

a. Adjusted easting values (X_1) used in the least-squares computation of coefficients. This is an independent variable.
b. Adjusted northing values (X_2) used in the least-squares computation of coefficients. This is an independent variable.
c. The dependent variable (Y) discussed in the text. In this example it was used to predict the X' pixel location.

TABLE 7–7 Coefficients used to rectify the Charleston, SC, Landsat Thematic Mapper scene.

$x' = -382.2366 + 0.034187x + (-0.005481)y$
$y' = 130{,}162 + (-0.005576)x + (-0.0349150)y$
where x, y are coordinates in the output image and x', y' are predicted image coordinates in the original, unrectified image.

X_2 = northing coordinate (y) of the map GCP

It is practical here to use X_1 and X_2 instead of x and y to simplify mathematical notation.

The seven coordinates used in the computation of the coefficients are shown in Table 7-6. Notice that the independent variables (X_1 and X_2) have been adjusted value = original value – minimum value) so that the sums of squares or sums of products do not become so large that they overwhelm the precision of the CPU being used. Note, for example, that most of the original northing UTM measurements associated with the map GCPs are already in the range of 3 million meters. The minimums subtracted in Table 7-6 are added back into the analysis at the final stage of coefficient computation. Next we will discuss the mathematics necessary to compute the x' coefficients shown in Table 7-7. It is technical, but should be of value to those interested in how to compute the coefficients used in Equation 7.9.

I. Find (X^TX) and (X^TY) in deviation form:

$n = 7$, the number of control points used

A. First compute:

$$\sum_{i=1}^{n} Y_i = 1{,}519 \qquad \sum_{i=1}^{n} X_{1i} = 24{,}165$$

$$\sum_{i=1}^{n} X_{2i} = 22{,}300 \qquad \bar{Y} = 217 \qquad \bar{X}_1 = 3{,}452.1428$$

$$\bar{X}_2 = 3{,}185.7142 \qquad \sum_{i=1}^{n} Y_i^2 = 366{,}491$$

$$\sum_{i=1}^{n} X_{1i}^2 = 119{,}151{,}925 \qquad \sum_{i=1}^{n} X_{2i}^2 = 89{,}832{,}800$$

$$\sum_{i=1}^{n} X_{1i}Y_i = 6{,}385{,}515 \qquad \sum_{i=1}^{n} X_{2i}Y_i = 5{,}234{,}140$$

$$\sum_{i=1}^{n} X_{1i}Y_{2i} = 91{,}550{,}500$$

B. Compute sums of squares:

1.

$$\sum_{i=1}^{n} X_{1i}^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_{1i}\right)^2 = 119{,}151{,}925 - \frac{1}{7}(24{,}165)^2$$

$$= 35{,}730{,}892.8571$$

2.

$$\sum_{i=1}^{n} X_{2i}^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_{2i}\right)^2 = 89{,}832{,}800 - \frac{1}{7}(22{,}300)^2$$

$$= 18{,}791{,}371.4286$$

3.

$$\sum_{i=1}^{n} X_{1i}X_{2i} - \frac{1}{n}\left(\sum_{i=1}^{n} X_{1i}\right)\left(\sum_{i=1}^{n} X_{2i}\right)$$

$$= 91{,}550{,}500 - \frac{1}{7}(24{,}165)(22{,}300)$$

$$= 14{,}567{,}714.2857$$

where

$$(X^TX) = \begin{Bmatrix} 35{,}730{,}892.8571 & 14{,}567{,}714.2857 \\ 14{,}567{,}714.2857 & 18{,}791{,}371.4286 \end{Bmatrix}$$

4. Covariance between Y and X_1:

$$\sum_{i=1}^{n} X_{1i}Y_i - \frac{1}{n}\left(\sum_{i=1}^{n} X_{1i}\right)\left(\sum_{i=1}^{n} Y_i\right)$$

$$= 6{,}385{,}515 - \frac{1}{7}(24{,}165)(1{,}519)$$

$$= 1{,}141{,}710$$

5. Covariance between Y and X_2:

$$\sum_{i=1}^{n} X_{2i}Y_i - \frac{1}{n}\left(\sum_{i=1}^{n} X_{2i}\right)\left(\sum_{i=1}^{n} Y_i\right)$$

$$= 5{,}234{,}140 - \frac{1}{7}(22{,}300)(1{,}519)$$

$$= 395{,}040$$

where

$$(X^TY) = \begin{Bmatrix} 1{,}141{,}710 \\ 395{,}040 \end{Bmatrix}$$

II. Find the inverse of $(X^TX) = (X^TX)^{-1}$:

A. First, find the determinant of the 2 ⊠ 2 matrix:

$$|X^TX| = (35{,}730{,}892.8571)(18{,}791{,}371.4286) - (14{,}567{,}714.2857)^2$$

$$= 459{,}214{,}179{,}643{,}488.9$$

B. Determine adjoint matrix of (X^TX) where adjoint equals the transpose of the cofactor matrix:

$$\text{Adjoint}^* = \begin{Bmatrix} 18{,}791{,}371.4286 & -14{,}567{,}714.2857 \\ -14{,}567{,}714.2857 & 35{,}730{,}892.8571 \end{Bmatrix}$$

*Note that if

$$A = \begin{Bmatrix} a & b \\ c & d \end{Bmatrix}, \text{ then } A^{-1} = \frac{1}{\det A}\begin{Bmatrix} d & -b \\ -c & a \end{Bmatrix}.$$

C. Get $(X^TX)^{-1}$ by multiplying the adjoint of (X^TX) by the det (X^TX) under 1.

$$(X^TX) = \left(\frac{1}{459{,}214{,}179{,}643{,}488.9}\right)$$

$$\begin{Bmatrix} 18{,}791{,}373.7142 & -14{,}567{,}714.2857 \\ -14{,}567{,}714.2857 & 35{,}730{,}896.4285 \end{Bmatrix}$$

$$= \begin{Bmatrix} 0.41\times10^{-7} & -0.32\times10^{-7} \\ -0.32\times10^{-7} & 0.78\times10^{-7} \end{Bmatrix}$$

III. Now, compute coefficients using

$$a_i = (X^TX)^{-1}(X^TY):$$

$$\begin{bmatrix} 0.41\times10^{7} & -0.32\times10^{-7} \\ -0.32\times10^{-7} & 0.78\times10^{-7} \end{bmatrix} \bullet \begin{bmatrix} 1{,}141{,}710 \\ 395{,}040 \end{bmatrix} =$$

$$a_1 = (0.000000041)(1{,}141{,}710) + (-0.000000032)(395{,}040)$$

$$a_2 = (-0.000000032)(1{,}141{,}710) + (-0.000000078)(395{,}040)$$

$$a_1 = 0.0341877$$

$$a_2 = -0.0054810$$

Now compute the intercept, a_0, from

$$a_0 = \bar{Y} - \sum_{i=1}^{2} a_i\bar{X}_i.$$

(The minimums of X_1 and X_2 [595,650 and 3,624,380, respectively] must be accounted for here. Please refer to Table 7-6.)

$$a_0 = 217 - [(0.0341877)(3{,}452.1428 + 595{,}650) +$$

$$(-0.005481)(3{,}185.7142 + 3{,}624{,}380)] = -382.2366479.$$

Therefore, the equation becomes

$$Y = -382.2366479 + 0.0341877X_1 - 0.0054810X_2.$$

Because we actually evaluated the dependent variable x', this becomes

$$x' = -382.2366479 + 0.0341877x - 0.0054810y$$

with x and y representing the map coordinates and x' being the predicted column coordinate in the original input image.

Similar procedures are required to compute the other three coefficients for the row (y') location in the input image. This would require inserting the seven y' pixel values in the equation in Table 7-6 instead of the x' pixel values just used. The X_1 and X_2 values associated with the x, y coordinates of the map remain the same.

Fill Output Matrix Using Spatial and Intensity Interpolation Resampling

With the coefficients computed, it was then necessary to 1) identify the UTM coordinates of the area on the map to be rectified, 2) select the type of intensity interpolation to be performed (e.g., nearest-neighbor, bilinear, or cubic convolution), and 3) specify the desired output pixel size. A nearest-neighbor resampling algorithm was selected with a desired output pixel dimension grid of 30 × 30 m. The finer the dimension of the output grid, the greater the number of computations required to fill it. Normally, the size of the pixel is made square (e.g., 30 × 30 m) to facilitate scaling considerations when the rectified data are displayed on computer monitors and various hard-copy output devices. These procedures resulted in rectified Landsat Thematic Mapper data.



Mosaicking

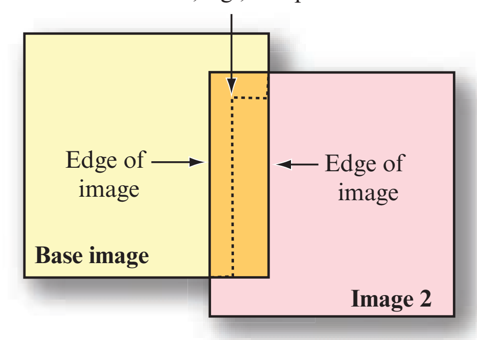
Mosaicking is the process of combining multiple images into a single seamless composite image. It is possible to mosaic unrectified individual frames or flightlines of remotely sensed data. However, it is more common to mosaic multiple images that have already been rectified to a standard map projection and datum (Figure 7-19).

Mosaicking Rectified Images

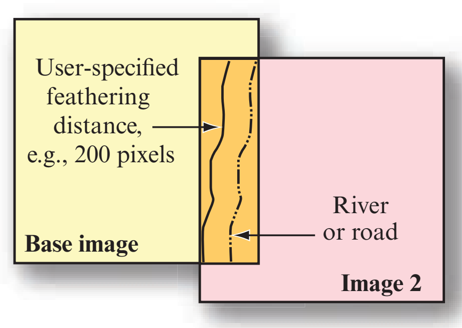
Mosaicking n rectified images requires several steps. First, the individual images should be rectified to the same map projection and datum. Ideally, rectification of the n images is performed using the same intensity interpolation resampling logic (e.g., nearest-neighbor) and pixel size (e.g., multiple Landsat TM scenes to be mosaicked are often resampled to 30 × 30 m).

Mosaic Feathering Logic

User-specified feathering distance, e.g., 200 pixels



a. Cut-line feathering.



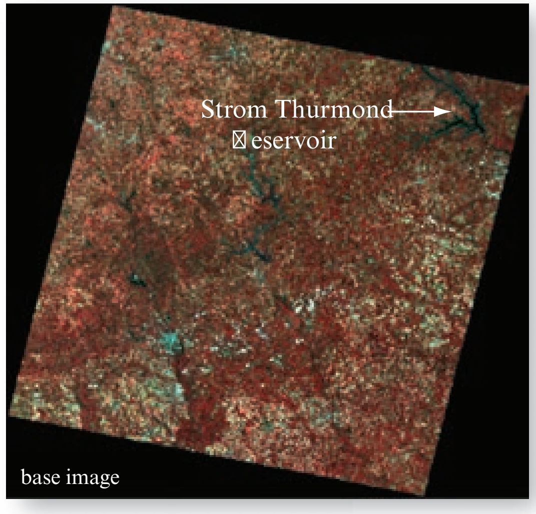
b. Edge feathering.

FIGURE 7–19 The visual seam between adjacent images being mosaicked may be minimized using a) cut-line feathering logic, or b) edge feathering.

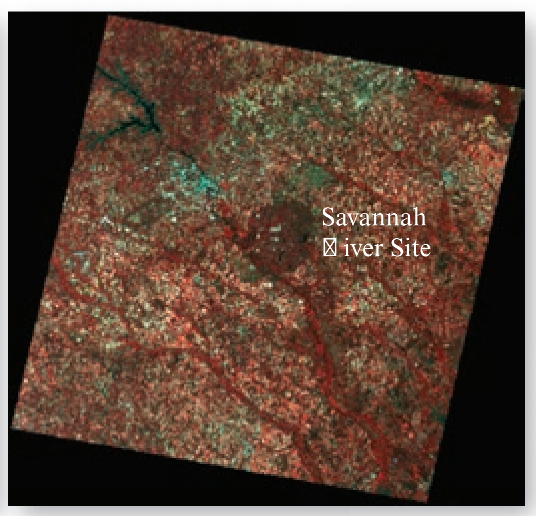
Next, one of the images to be mosaicked is designated as the *base image*. The base image and image 2 will normally overlap a certain amount (e.g., 20% – 30%). A representative geographic area in the overlap region is identified. This area in the base image is contrast stretched according to user specifications. The histogram of this geographic area in the base image is extracted. The histogram from the base image is then applied to image 2 using a histogram-matching algorithm. This causes the two images to have approximately the same brightness value characteristics.

It is possible to have the pixel brightness values in one scene simply dominate the pixel values in the overlapping scene. Unfortunately, this can result in noticeable seams in the final mosaic. Therefore, it is common to blend the seams between mosaicked images using *feathering* (Tucker et al., 2004). Some digital image processing systems allow the user to specific a feathering buffer distance (e.g., 200 pixels) wherein 0% of the

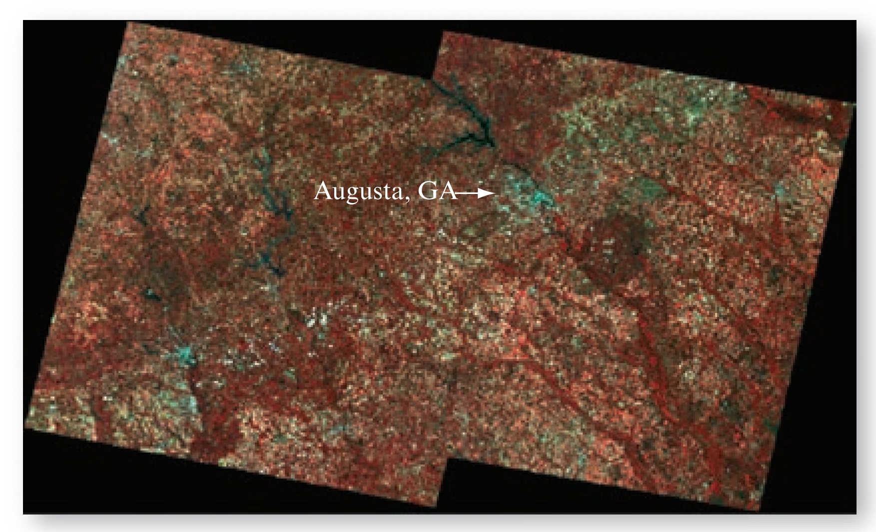
Image Mosaicking



a. Rectified Landsat ETM+ image of eastern Georgia obtained on October 3, 2001 (bands 4, 3, 2; Worldwide Reference System—Path 18, Row 37).



b. Rectified Landsat ETM+ image of western South Carolina obtained on October 26, 2001 (bands 4, 3, 2; Worldwide Reference System—Path 17, Row 37).



c. Feathered mosaic of rectified Landsat ETM+ imagery of eastern Georgia and western South Carolina.

FIGURE 7–20 Two Landsat Enhanced Thematic Mapper Plus (ETM^+) images are mosaicked using feathering logic. Original imagery courtesy of NASA.

base image is used in the blending at the edge and 100% of image 2 is used to make the output image (e.g., Figure 7-19a). At the specified distance (e.g., 200 pixels) in from the edge, 100% of the base image is used to make the output image and 0% of image 2 is used. At 100 pixels in from the edge, 50% of each image is used to make the output file.

Sometimes analysts prefer to use a linear feature such as a river, road, or agricultural field boundaries to subdue the edge between adjacent mosaicked images. In this case, the analyst identifies a polyline in the image (using an annotation tool) and then specifies a buffer distance away from the line as before where the feathering will take place (Figure 7-19b). It is not absolutely necessary to use natural or human-made features when performing cut-line feathering. Any user-specified polyline will do. The output file is then produced. It consists of two histogram-matched images that have been feathered at the common edge(s) of the dataset. Hopefully, the multiple image edge-match is not visible. Additional images (3, 4, 5, etc.) to be mosaicked are histogram-matched and feathered using similar logic.

The output file of n mosaicked images should appear as if it were one continuous image. For example, Figure 7-20a–b depicts two rectified Landsat Enhanced Thematic Mapper Plus (ETM^+) scenes obtained over eastern Georgia and western South Carolina on October 3,

2011 Mosaic of LANDSAT EMT$^+$ Imagery of the Conterminous United States

a. Annual.

WELD: Web-enabled Landsat Data

b. Annual and seasonal.

c. Monthly.

d. Weekly.

FIGURE 7–21 a) 2011 annual mosaic of Landsat ETM$^+$ image data. ETM$^+$ imagery was provided by the USGS (http://weld.cr.usgs.gov/region_ds.php). Digital image processing was performed by the NASA-sponsored *WELD: Web-enabled Landsat Data* project (Roy et al., 2010; 2011; USGS WELD, 2014). b) Annual and seasonal periods of interest. c) Monthly periods of interest. d) Weekly periods of interest. WELD interface and imagery courtesy of the U.S. Geological Survey and NASA.

2001, and October 26, 2001, respectively. Each image was resampled using nearest-neighbor logic to 30×30 m pixels in a UTM projection. The two Landsat ETM$^+$ frames contained approximately 20% side-lap. Note the Strom Thurmond Reservoir in the overlap region. The two images were histogram-matched and mosaicked using edge-feathering logic (Figure 7-20c). This is a good mosaic because it is difficult to detect the seam between the two input images.

Mosaicked images that span entire counties, states, and countries are especially important for many biological

Mosaic of Three Rectified NASA ATLAS Multispectral Flightlines of Myrtle Beach, SC

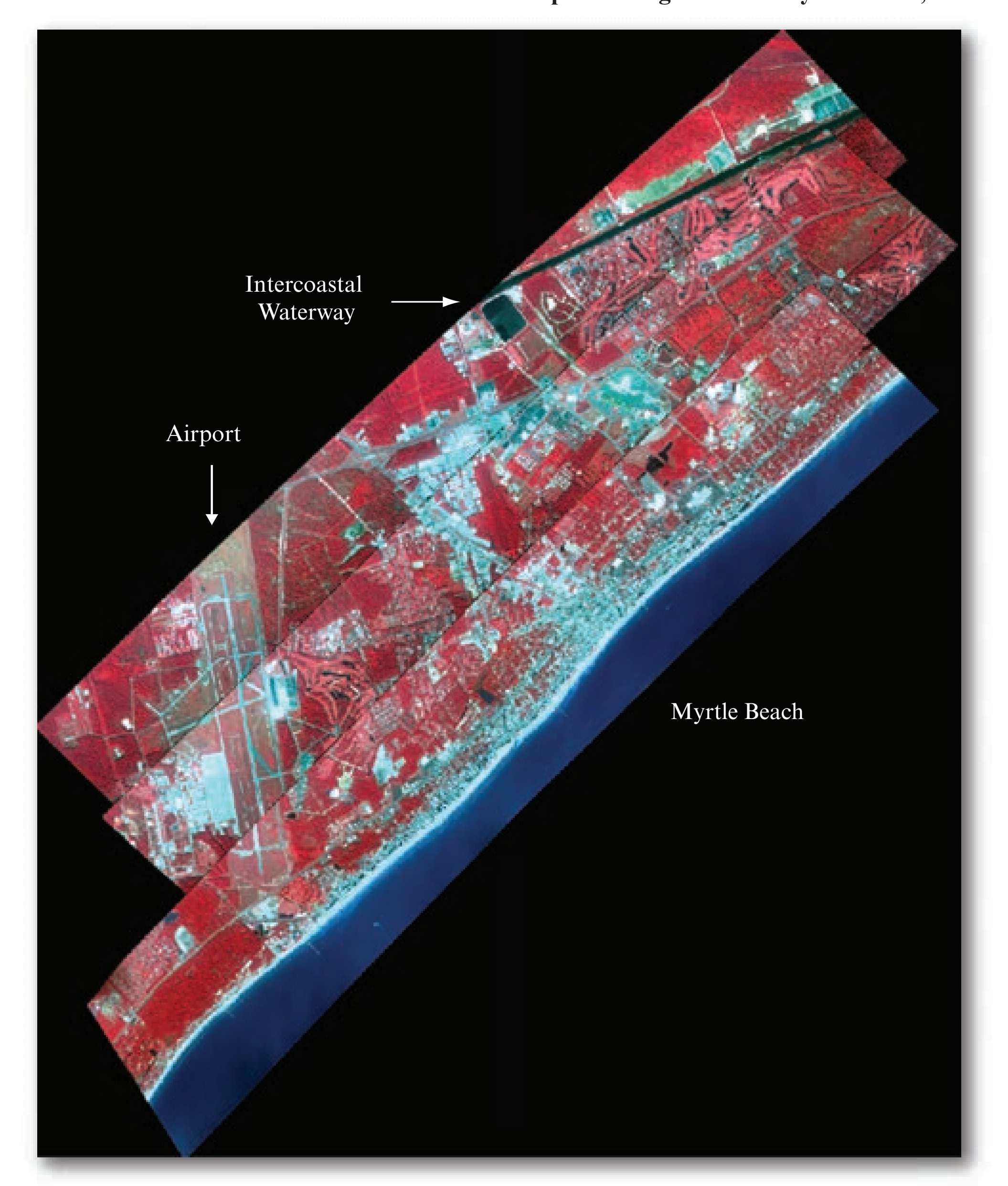
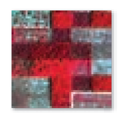


FIGURE 7–22 Mosaic of three NASA ATLAS 1 × 1 m multispectral flightlines of Myrtle Beach, SC, collected on October 7, 1997 (original imagery courtesy of NASA).

and geophysical investigations. For this reason, numerous agencies and data centers routinely provide mosaicked products for scientific research. For example, the NASA-funded *Web-enabled Landsat Data (WELD)* project systematically generated 30 × 30 m composite Landsat Enhanced Thematic Mapper Plus (ETM^+) mosaics of the conterminous United States and Alaska from 2002 to 2012 (USGS WELD, 2014). A much-reduced Landsat ETM^+ mosaic of the conterminous United States in 2011 is shown in Figure 7-21a. It is composed of hundreds of ETM^+ images acquired during 2011. The mosaicking methodology is described in Roy et al. (2010; 2011). Unlike most mosaicking projects, WELD does not just select cloud-free scenes to be included in the mosaic. Instead, it analyzes a number of scenes in the annual/seasonal, monthly, or weekly time period under investigation and uses special criteria to create a composite mosaic on a *pixel-by-pixel* basis. The criteria (e.g., per pixel saturation, cloudiness, NDVI <0.5, maximal NDVI) are specified in the *WELD Algorithm Theoretical Basis Document* (Roy et al., 2011). The mosaic output is in an Albers Equal Area Projection, which is ideal for east-west extending areas such as the conterminous United States (Snyder, 1995).

Mosaicking multiple flightlines of aircraft panchromatic, multispectral, or hyperspectral is usually more

difficult than rectifying satellite imagery. A mosaic of three flightlines of NASA ATLAS 1 × 1 m multispectral data is shown in Figure 7-22. The color balance between the three flightlines is reasonable but there is some noticeable geometric error at the common edges of the flightlines. Additional carefully selected ground control points and perhaps a higher order polynomial could improve the mosaic.



Conclusion

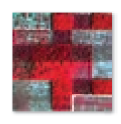
Toutin (2004) posed the question "Why does the geometric correction process seem to be more important today than before?" He suggests that in 1972, the importance of image geometric distortion was negligible because:

- Landsat MSS-like images were acquired at Nadir at relatively coarse spatial resolution, typically > 80 × 80 m;
- the images were in hard-copy paper or positive transparency format and analyzed visually; and
- there was very little fusion of multiple types of geospatial information taking place.

Today, geometric distortion is much more of a concern because:

- the spatial resolution of the remote sensor data is much finer (e.g., often < 1 × 1 m)
- many images are acquired at off-Nadir viewing angles [Chiu et al., 2011]);
- the image data and the remote sensing–derived output products are digital (Devaraj and Shah, 2014);
- it is common to fuse or merge different types of remote sensor data to complete a project (e.g., Green, 2011); and
- it is common to merge both raster and vector digital geospatial data for modeling purposes.

Geometrically accurate remote sensor data are required to create remote sensing-derived products that the general public can trust.



References

Bossler, J. D. (Ed.), 2010, *Manual of Geospatial Science & Technology*, 2nd Ed., London: Taylor & Francis, 832 p.

Brown, L. G., 1992, "A Survey of Image Registration Techniques," *ACM Computing Surveys*, 24(4):325–376.

Buiten, H. J., and B. Van Putten, 1997, "Quality Assessment of Remote Sensing Registration—Analysis and Testing of Control Point Residuals," *ISPRS Journal of Photogrammetry & Remote Sensing*, 52:57–73.

Chiu, L. S., Hao, X., Resmini, R. G., Sun, D., Stefanidis, A., Qu, J. J., and R. Yang, 2011, "Chapter 2: Earth Observations," in Yang, C., Wong, D., Miao, Q., and R. Yang (Eds.) *Advanced Geoformation Science*, Boca Raton: CRC Press, 485.

Devaraj, C., and C. A. Shah, 2014, "Automated Geometric Correction of Multispectral Images from High Resolution CCD Camera (HRCC) on-board CBERS-2 and CBERS-2B," *ISPRS Journal of Photogrammetry & Remote Sensing*, 89:13–24.

Esri, 2004, *Understanding Map Projections*, Redlands: Esri, Inc., 120 p.

Furuti, C. A., 2011, *Map Projections*, (http://www.progonos.com/furuti).

Garnett, W., 2009, *A Little Book on Map Projection*, London: General Books, 62 p.

Gibson, P. J., and C. H. Power, 2000, *Introductory Remote Sensing: Digital Image Processing and Applications*, NY: Routledge, 249 p.

Gonzalez, R. C., and R. E. Woods, 2007, *Digital Image Processing*, 3rd Ed., NY: Addison-Wesley, 797 p.

Grafarend, E. W., and Krumm, F. W., 2006, *Map Projections: Cartographic Information Systems*, NY: Springer, 714 p.

Green, K., 2011, "Change Matters," *Photogrammetric Engineering & Remote Sensing*, 77(4):305–309.

Iliffe, J.C., 2008, *Datums and Map Projections for Remote Sensing, GIS, and Surveying*, 2nd Ed., NY: Whittles, 208 p.

Im, J., Jensen, J. R., Coleman, M., and E. Nelson, 2009, "Hyperspectral Remote Sensing Analysis of Short Rotation Woody Crops Grown with Controlled Nutrient and Irrigation Treatments," *Geocarto International*, 24(4):293–312.

Intergraph, 2013, *ERDAS Field Guide*, Huntsville: Intergraph, 772 p.

Jensen, J. R., Botchway, K., Brennan-Galvin, E., Johannsen, C., Juma, C., Mabogunje, A., Miller, R., Price, K., Reining, P., Skole, D., Stancioff, A., and D. R. F. Taylor, 2002, *Down to Earth: Geographic Information for Sustainable Development in Africa*, Washington: National Research Council, 155 p.

Jensen, J. R. and R. R. Jensen, 2013, *Introductory Geographic Information Systems*, Boston: Pearson, 400 p.

Jensen, J. R., Cowen, D., Narumalani, W., Weatherbee, O., and J. Althausen, 1993, "Evaluation of CoastWatch Change Detection Protocol in South Carolina," *Photogrammetric Engineering & Remote Sensing*, 59(6):1039–1046.

Kanters, F., 2007, *Small-scale Map Projection Design*, London: Taylor & Francis, 352 p.

Konecny, G., 2014, *Geoinformation: Remote Sensing, Photogrammetry and Geographic Information Systems*, Boca Raton: CRC Press, 416 p.

Krygier, J. B., 2011, Course on *Cartography and Visualization*, Delaware, OH: Department of Geology & Geography, Ohio Wesleyan University.

Lillesand, T., Keifer, R., and J. Chipman, 2008, *Remote Sensing and Image Interpretation*, 6th Ed., NY: John Wiley & Sons, 756 p.

Maher, M. M., 2010, *Lining Up Data in ArcGIS: A Guide to Map Projections*, Redlands: Esri Press, 200 p.

Marcomini, A., Suter, G. W., and A. Critto, 2008, *Decision Support Systems for Risk-based Management of Contaminated Sites*, New York: Springer, 435 p.

McRoberts, R. E., 2010, "The Effects of Rectification and Global Positioning System Errors on Satellite Image-based Estimates of Forest Area," *Remote Sensing of Environment*, 114:1710–1717.

Merchant, J. W. and S. Narumalani, 2009, "Integrating Remote Sensing and Geographic Information Systems," Chapter 18 in *The Sage Handbook of Remote Sensing*, Warner, T., Nellis, M. D. and J. Foody (Eds.), Los Angeles: Sage Publications, 257–268.

NASA, 1998, *Landsat 7 Initial Assessment Geometric Algorithm Theoretical Basis Document*, Washington: NASA, 177 p.

NASA, 2015, *G.Projector*, New York: NASA Goddard Institute for Space Studies. *G.Projector* was written by R. B. Schmunk. Software can be downloaded from http://www.giss.nasa.gov/tools/gprojector/.

NGA, 2007, *The Universal Grid System: A Simplified Definition and Explanation of UTM and Related Systems,* Washington: National Geospatial-Intelligence Agency Office of GEOINT Sciences, 8 p.

Niblack, W., 1986, *An Introduction to Digital Image Processing*, Englewood Cliffs, NJ: Prentice-Hall, 215 p.

Novak, K., 1992, "Rectification of Digital Imagery," *Photogrammetric Engineering & Remote Sensing*, 58(3):339–344.

Purkis, S. and V. Klemas, 2011, *Remote Sensing and Global Environmental Change*, NY: John Wiley, 367 p.

Richards, J. A., 2013, *Remote Sensing Digital Image Analysis*, 5th Ed., New York: Springer-Verlag, 494 p.

Robinson, A. and J. P. Snyder, 1991, *Matching the Map Projection to the Need,* Bethesda: American Congress on Surveying and Mapping, 30 p.

Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M.C., Loveland, T.R., Vermote, E.F., Zhang, C., 2010, "Web-enabled Landsat Data (WELD): Landsat ETM^+ Composited Mosaics of the Conterminous United States," *Remote Sensing of Environment*, 114: 35–49.

Roy, D. P., Ju, J., Kommareddy, I., Hansen, M., Vermote, E., Zhang, and C., Kommareddy, A., 2011, *Web Enabled Landsat Data (WELD) Products—Algorithm Theoretical Basis Document*, February, http://globalmonitoring.sdstate.edu/projects/weld/WELD_ATBD.pdf.

Schowengerdt, R. A., 2007, *Remote Sensing: Models and Methods for Image Processing*, 3rd Ed., San Diego: Academic Press, 515 p.

Slocum, T. A., McMaster, R. B., Kessler, F. C. and H. H. Howard, 2008, *Thematic Cartography and Geographic Visualization*, 3rd Ed., Boston: Pearson, 520 p.

Snyder, J. P., 1987, *Map Projections: A Working Manual*, U.S. Geological Survey Professional Paper #1395, Washington: U.S. Government Printing Office.

Snyder, J. P., 1995, *Flattening the Earth: Two Thousand Years of Map Projections*, Chicago: University of Chicago Press.

Sugumaran, S. and J. DeGrotte, 2010, *Spatial Decision Support Systems: Principles and Practices*, Boca Raton: CRC Press, 486 p.

Toutin, T., 2004, Review Article: Geometric Processing of Remote Sensing Images: Models, Algorithms and Methods, *International Journal of Remote Sensing*, 25(10):1893–1924.

Tucker, C. J., Grant, D. M., and J. Dykstra, 2004, "NASA's Global Orthorectified Landsat Data Set," *Photogrammetric Engineering & Remote Sensing*, 70(3):313–322.

USGS, 2011, *Map Projections—A Brochure*, Washington: U.S. Geological Survey.

USGS WELD, 2014, *Web-enabled Landsat Data (Weld) Projects*, NASA: Washington, https://landsat.usgs.gov/WELD.php.

Van der Meer, F., van der Werff, H., and S. M. de Jong, 2009, "Chapter 16: Pre-Processing of Optical Imagery," in Warner, T. A., Nellis, M. D. and G. M. Foody (Eds.), *The Sage Handbook of Remote Sensing*, Los Angeles: Sage Publications, 229–243.

Williams, D., 2003, Correspondence regarding the angular velocity of Landsat satellites 1 to 5 and 7, Greenbelt: NASA Goddard Space Flight Center.

Wolberg, G., 1990, *Digital Image Warping*, NY: John Wiley–IEEE Computer Society, 340 p.

Wolf, P. R., Dewitt, R. A., and B. E. Wilkinson, 2013, *Elements of Photogrammetry with Applications in GIS*, 4th Ed., NY: McGraw-Hill, 640 p.

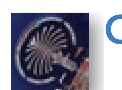
8 IMAGE ENHANCEMENT





Source: NASA

Enhancement algorithms are applied to remotely sensed data to improve the appearance of an image for human visual analysis or occasionally for subsequent machine analysis. There is no such thing as the ideal or best image enhancement because the results are ultimately evaluated by humans, who make subjective judgments as to whether an enhanced image is useful.



Overview

This chapter identifies a variety of digital image enhancement operations that have proven of value for visual analysis of remote sensor data and/or subsequent machine analysis. **Point operations** modify the brightness values of each pixel in an image dataset independent of the characteristics of neighboring pixels. **Local operations** modify the value of each pixel in the context of the brightness values of the pixels surrounding it. Various types of image enhancement using either point or local operations are introduced including: image reduction and magnification, the collection of information along a transect (a spatial profile), contrast stretching, density slicing, spatial filtering, Fourier analysis, vegetation transforms, texture mapping, and the derivation of several landscape ecology metrics.

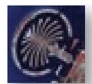


Image Reduction and Magnification

Image analysts routinely view images that have been reduced in size or magnified during the image interpretation process. Image reduction techniques allow the analyst to obtain a regional perspective of the remotely sensed data. Image magnification techniques allow the analyst to zoom in and view very site-specific pixel characteristics.

Image Reduction

In the early stages of a remote sensing project it is often necessary to view the entire image in order to locate the row and column coordinates of a subimage that encompasses the study area. Most commercially available remote sensor data are composed of more than 3,000 rows × 3,000 columns and a number of bands. Unfortunately, most digital image processing system displays cannot display this many pixels at one time. Therefore, it is useful to have a simple procedure for reducing the size of the original image dataset down to a smaller dataset that can be viewed on the screen at one time for orientation purposes. To *reduce* a digital image to just $1/m^2$ of the original data, every mth row and mth column of the imagery are systematically selected and displayed. For example, consider the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) image of Oahu, HI, that originally had 4,104 rows × 3,638 columns shown in Figure 8-1. The size of the original image was reduced by using every other row and every other column (i.e., $m = 2$) in the dataset. The sampled image shown consists of just 2,052 rows × 1,819 columns. This reduced dataset contains only one fourth (25%) of the pixels found in the original scene. The logic associated with a simple 2× integer reduction is shown in Figure 8-2a.

If you were to compare the original ASTER data with the reduced ASTER data, there is an obvious loss of detail because so many of the pixels are not present. Therefore, we rarely apply digital image processing techniques to image reductions. Instead, they are used

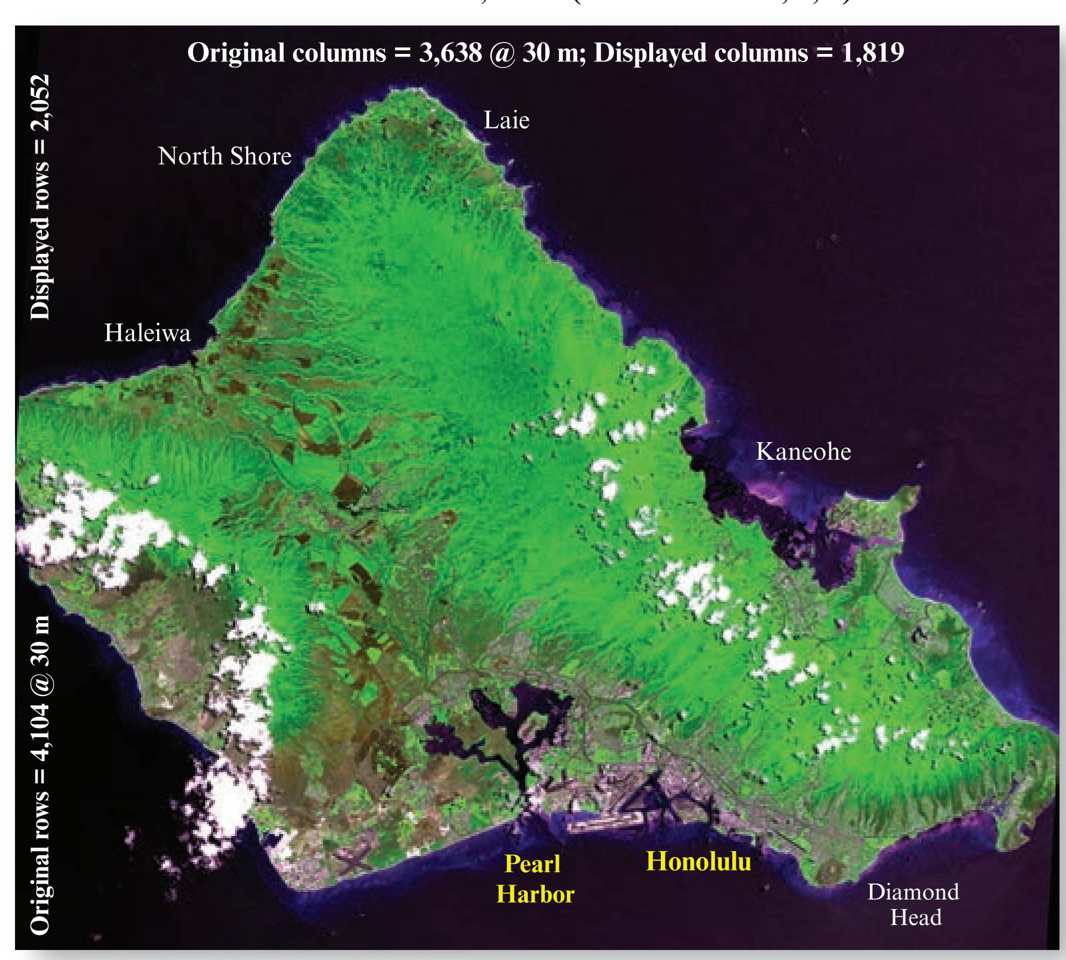


FIGURE 8–1 This 2,052 row × 1,819 column image represents only 25% of the data found in the original 4,104 row × 3,638 column NASA Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) image. It was created by sampling every other row and column using the logic shown in Figure 8-2 (courtesy NASA/GSFC/METI/ERSDAC/JAROS and U.S./Japan ASTER Science Team).

for orienting within a scene and locating the row and column coordinates of specific areas-of-interest (AOI) that can then be extracted at full resolution for image analysis.

Image Magnification

Digital image *magnification* (often referred to as *zooming in*) is usually performed to enlarge the scale of an image or map for visual interpretation or, occasionally, to match the scale of another image or map. Just as row and column deletion is the simplest form of image or map reduction, row and column replication represents the simplest form of digital magnification. To magnify a digital image or map by an integer factor $m\times$, each pixel in the original image or map is usually replaced by an $m \times m$ block of pixels, all with the same value as the original input pixel. The logic of a 2× magnification is shown in Figure 8-2b. This form of magnification is characterized by visible square tiles of pixels in the output display. Image magnifications of 1×, 2×, and 3× applied to the ASTER image of Oahu, HI, are shown in Figure 8-3.

Most sophisticated digital image processing systems allow an analyst to specify floating-point magnification (or reduction) factors (e.g., zoom in 2.75×). This requires that the original remote sensor or map data be resampled in near real-time using one of the standard resampling algorithms (e.g., nearest-neighbor, bilinear interpolation, or cubic convolution discussed in Chapter 7). This is a very useful technique when the analyst is trying to obtain detailed information about the characteristics of a small geographic area of interest. During the training phase of a supervised classification (to be discussed in Chapter 9), it is especially useful to be

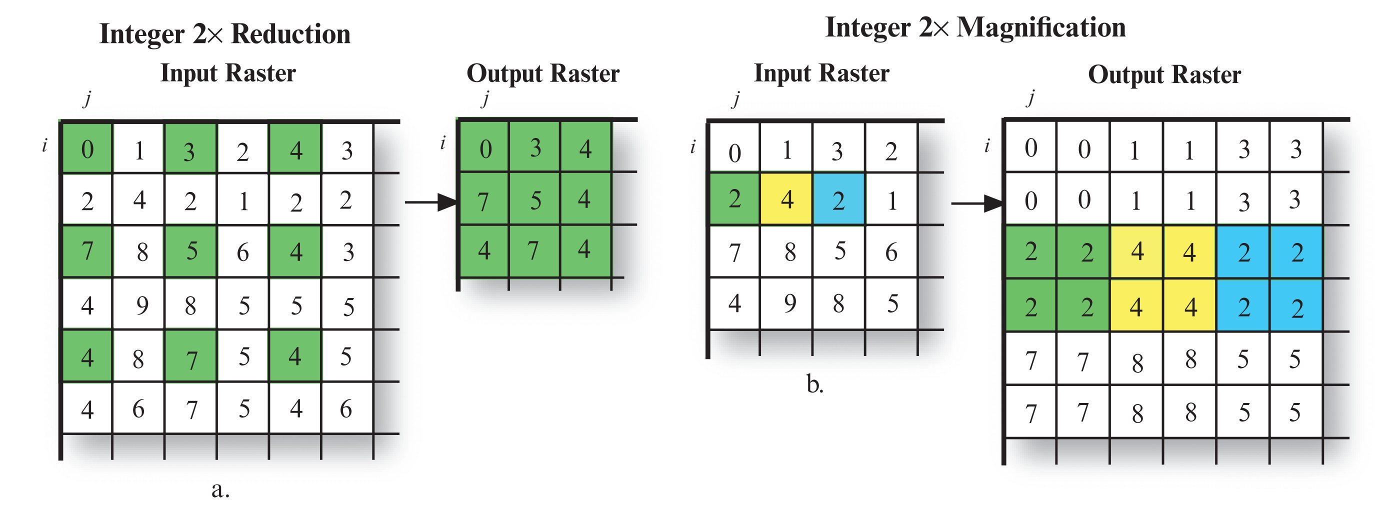
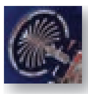


FIGURE 8–2 a) The logic associated with 2× image integer reduction using local pixel operations that sample every other row and column of the original dataset. This operation results in a new image consisting of only one quarter (25%) of the original data. b) The logic of 2× image magnification.

able to zoom in to the raw remote sensor data at very precise floating point increments to isolate a particular field or body of water.

In addition to magnification, virtually all digital image processing systems provide a mechanism whereby the analyst can *pan* or *roam* about a much larger geographic area (e.g., 2,048 × 2,048) while viewing only a portion (e.g., 512 × 512) of this area at any one time. This allows the analyst to view parts of the database much more rapidly.



Transects (Spatial Profiles)

The ability to extract brightness values along a user-specified **transect** (also referred to as a **spatial profile**) between two or multiple points in a single-band or multiple-band color composite image is important in many remote sensing image interpretation applications. For example, consider the three simple 50-pixel horizontal spatial profiles (A, B, and C) identified in the single-band, black-and-white thermal infrared image of the Savannah River (Figure 8-4a). In each case, the spatial profile in histogram format depicts the magnitude of the brightness value at each pixel along the 50-pixel transect. Each pixel in the thermal image was 2.8 × 2.8 m.

Sometimes it is useful to view the spatial profile histogram in a single grayscale tone (or color) as shown in Figure 8-4b. Conversely, analysts often prefer to display the grayscale tone of the individual brightness values encountered along the transect as shown in Figures 8-4c, d, and e.

Each of the transects in Figure 8-4 was located along a single horizontal scan line. In this case, each of the transects was exactly 140 m long (50 pixels × 2.8 m = 140 m). But what if we wanted to extract a detailed transect across the river perpendicular to the shoreline? Or what if we wanted to run a transect through the heart of the plume to see how fast the thermal plume was cooling after entering the river? In this case it would be necessary to take into account the stair-stepped nature of the transect and incorporate distance information derived using the Pythagorean theorem previously discussed. Examples of a 7-pixel diagonal transect and a 7-pixel horizontal transect are shown in Figure 8-5.

Accurate transect results can also be obtained by rotating the image until the desired transect is aligned with either a single line or column in the dataset. For example, consider the density-sliced image of the Savannah River thermal plume shown in Figure 8-6b based on the color look-up table values in Table 8-1. In this example, we are interested in identifying a) the temperature of the thermal plume in the river perpendicular to the shoreline using a 50-pixel transect (1), and b) the rate of temperature decrease as the plume progresses downstream using a 100-pixel transect (2). The brightness values encountered along the two transects were extracted only after the original image was geometrically rotated 16° in the clockwise direction so that the endpoints of each transect fell on the same scan line or

Magnification of ASTER Imagery of Oahu, HI



a. 1× magnification (original resolution).



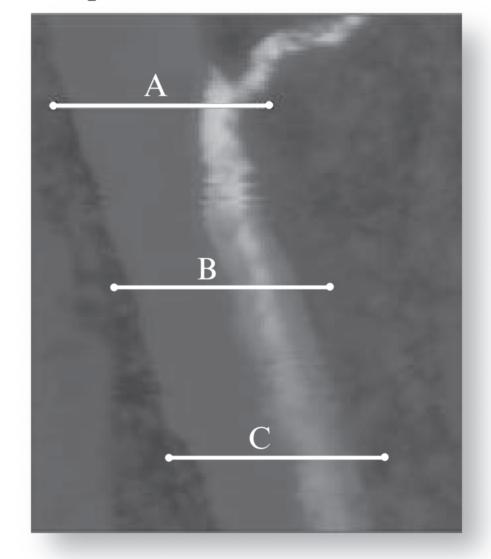
b. 2× magnification.



c. 3× magnification.

FIGURE 8–3 NASA ASTER imagery of Oahu, HI, magnified 1×, 2×, and 3× (courtesy NASA/GSFC/METI/ERSDAC/JAROS and U.S./Japan ASTER Science Team).

Spatial Profiles - Transects



a. Predawn thermal infrared imagery.

b. Transect A' — Brightness Value vs. Distance in Pixels

c. Transect A — Brightness Value vs. Distance in Pixels

d. Transect B — Brightness Value vs. Distance in Pixels

e. Transect C — Brightness Value vs. Distance in Pixels

FIGURE 8–4 a) Three 50-pixel spatial profiles (transects) passed through the Savannah River predawn thermal infrared data. Each pixel in the dataset is 2.8 × 2.8 m. b) The spatial profile data are displayed in histogram format using a single grayscale tone. c–e) The spatial profile data are displayed in histogram format according to their original brightness values.

column (Figure 8-6c). This ensured that the number of meters in each temperature class along each transect was accurately measured. If the analyst extracts transects where the endpoints do not fall on the same scan line (or column), the hypotenuse of stair-stepped pixels must be considered instead of the simple horizontal pixel distance (Figure 8-5).

Transects 1 and 2 are shown in Figure 8-6d and e. The relationship between the original brightness values and the class intervals of the transects is provided in Table 8-1. By counting the number of pixels along a transect in specific temperature-class intervals within the plume and counting the total number of pixels of river (Table 8-1), it is possible to determine the proportion of the thermal plume falling within specific temperature-class

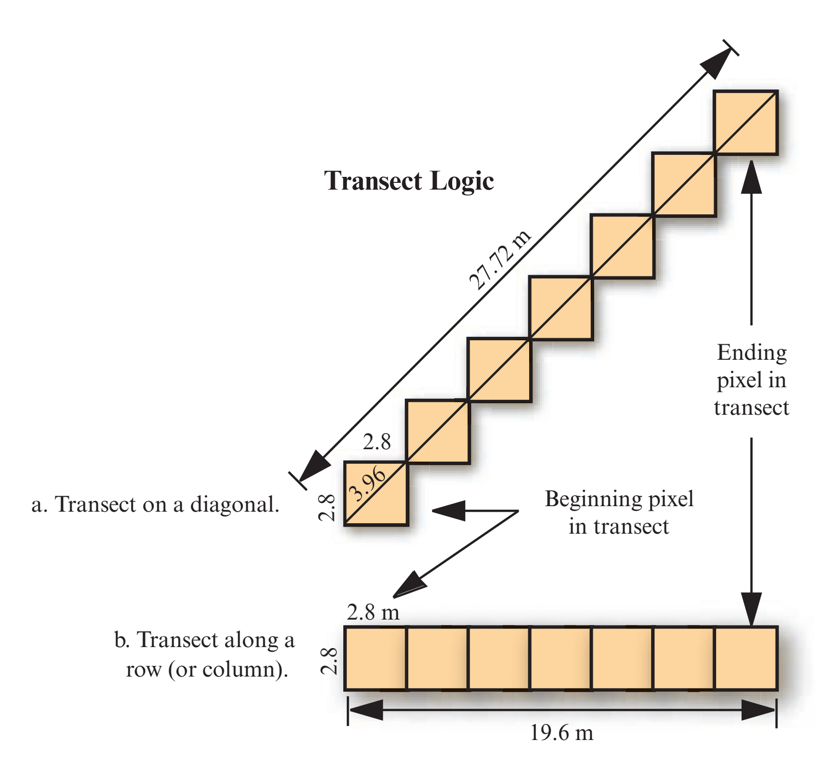


FIGURE 8–5 a) The stair-stepped nature of a transect (spatial profile) when the beginning and ending points are not on the same line (or column). In this example, a 45° angle transect consisting of just seven 2.8 × 2.8 m pixels would be 27.72 m in length. b) A 7-pixel transect along a row (or column) of the image would be only 19.6 m long. Therefore, when diagonal transects (or spatial profiles) are extracted, it is necessary to compute their length based on the Pythagorean theorem.

TABLE 8–1 Savannah River thermal plume density slice specifications used to create Figure 8-6bc. The number of pixels in each class interval associated with **Transect 1** is provided.

		Relationship of Class to Ambient River Temperature						
		Class 1 Dark blue (RGB = 0, 0, 120) Ambient	Class 2 Light blue (RGB = 0, 0, 255) +1°C	Class 3 Green (RGB = 0, 255, 0) 1.2°–2.8°C	Class 4 Yellow (RGB = 255, 255, 0) 3.0°–5.0°C	Class 5 Orange (RGB = 255, 50, 0) 5.2°–10°C	Class 6 Red (RGB = 255, 0, 0) 10.2°–20°C	Class 7 White (RGB = 255, 255, 255) >20°C
Transect 1[a]	**Width of River**[b]	**Brightness Value Range for Each Class Interval**						
		74–76	77–80	81–89	90–100	101–125	126–176	177–255
50 pixels @ 2.8 m	39 pixels = 109.2 m	24/67.2[c]	2/5.6	1/2.8	1/2.8	2/5.6	5/14	4/11.2

a. Transect 1 was 140 m long (50 pixels at 2.8 m/pixel). Transect measurements in the river were made only after the image was rotated so that the beginning and ending pixels of the transect fell on the same scan line.
b. Includes 1 mixed pixel of land and water on each side of the river.
c. Notation represents pixels and meters; for example, 24 pixels represent 67.2 m.

intervals (Jensen et al., 1983; 1986). In 1981, the South Carolina Department of Health and Environmental Control (DHEC) mandated that a thermal plume could not be >2.8° above river ambient temperature for more than one third of the width of the river. Transect 1 information extracted from thermal infrared imagery (summarized in Table 8-1) can be used to determine if the plume was in compliance at that location.

FIGURE 8–6 Transects (spatial profiles) passed through a predawn thermal infrared image of a thermal plume located in the Savannah River. a) Original image. b) Original image density-sliced according to the logic presented in Table 8-1. c) Density-sliced image rotated 16° with Transects 1 and 2 displayed. d) Spatial profile of Transect 1. e) Spatial profile of Transect 2.



Spectral Profiles

In addition to extracting pixel brightness values or percent reflectance values along selected spatial profiles (transects), it is often useful to extract the full spectrum of brightness values in n bands for an individual pixel. This is commonly referred to as a **spectral profile**.

In a spectral profile, the x-axis identifies the number of the individual bands in the dataset and the y-axis documents the brightness value (or percent reflectance if the data have been calibrated) of the pixel under investigation for each of the bands (Figure 8-7e).

The usefulness of the spectral profile is dependent upon the quality of information in the spectral data. Analysts sometimes assume that in order to do quality

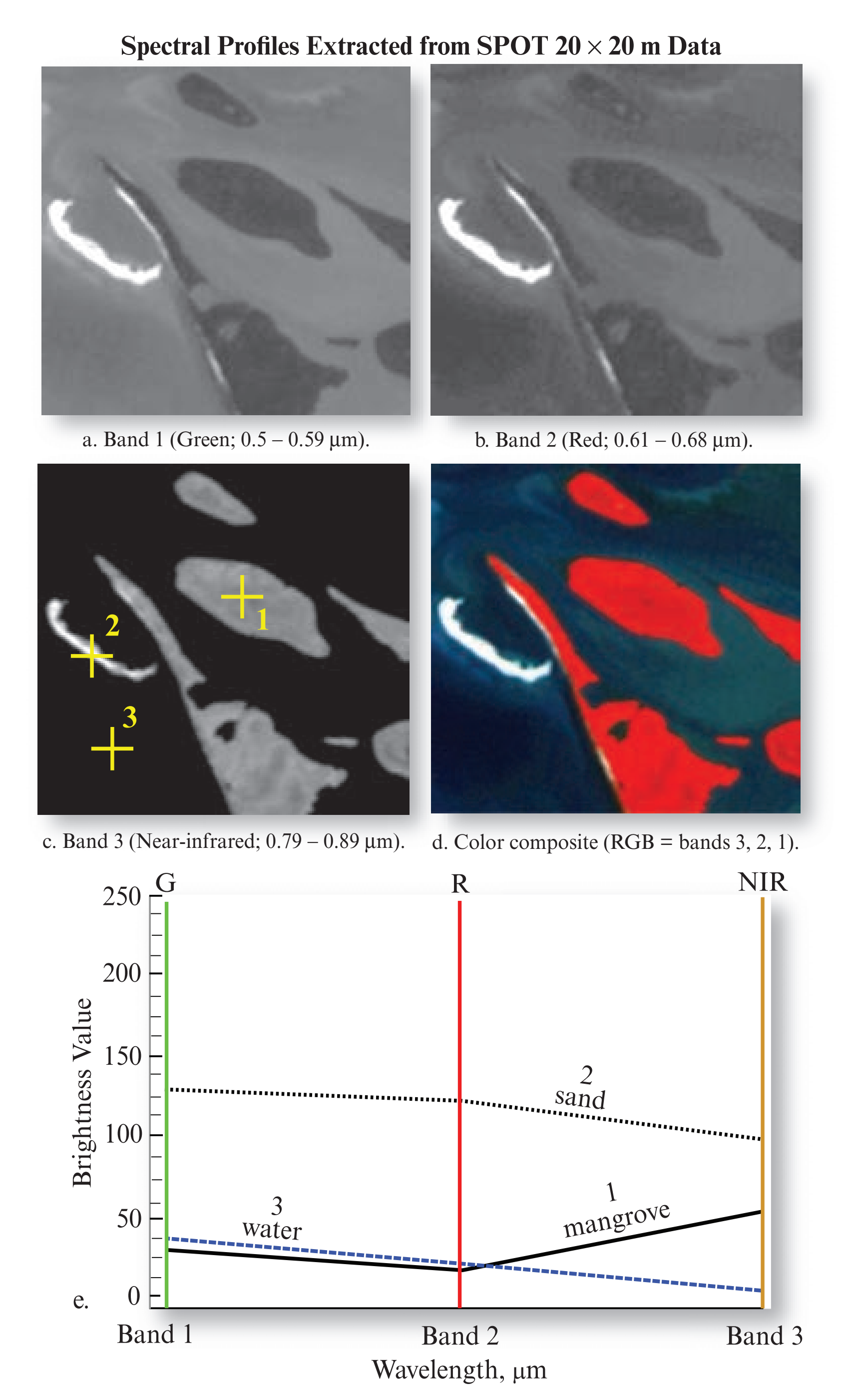


FIGURE 8–7 a–c) Three bands of SPOT 20 × 20 m multispectral data of Marco Island, FL. d) A color composite of SPOT bands 3, 2, and 1. e) Spectral profiles of mangrove, beach sand, and ocean water extracted from the multispectral data (SPOT imagery courtesy of Airbus Defense and Space).

remote sensing research they need a tremendous number of bands. Sometimes this is true. However, at other times just two or three optimally located bands in the electromagnetic spectrum can be sufficient to extract the desired information and solve a problem. Therefore, the goal is to have just the right number of optimally located, non-redundant spectral bands. Spectral profiles can assist the analyst by providing unique visual and quantitative information about the spectral characteristics of the phenomena under investigation and whether there are any serious problems with the spectral characteristics of the dataset.

Spectral profiles extracted from 1) mangrove, 2) sand, and 3) water locations in a 20 × 20 m three-band SPOT image (green, red, and near-infrared) of Marco Island, FL, are shown in Figure 8-7e. Note that these data have not been converted into percent reflectance;

Spectral Profiles Extracted from HyMap Hyperspectral Data

a. Band 9 (Green; 0.5591 μm).

b. Band 15 (Red; 0.6508 μm).

c. Band 40 (Near-infrared; 1.0172 μm).

d. Color composite (RGB = bands 40, 15, 9).

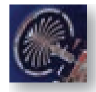
e.

FIGURE 8–8 a–c) Three bands of HyMap hyperspectral data of the Debordiu colony near North Inlet, SC. The data were obtained at a spatial resolution of 3 × 3 m. d) Color composite of HyMap bands 40, 15, and 9. e) Spectral profiles of golf "green" turf, sand, roof, and water extracted from the 116 bands of hyperspectral data.

therefore, the *y*-axis is simply labeled Brightness Value. Because this dataset consists of only three bands, it is useful but not very informative. As expected, the pixel of mangrove under investigation (1) absorbs more red light than green light due to chlorophyll *a* absorption and reflects significant amounts of the incident near-infrared energy. The sandy beach (2) reflects approximately equal amounts of green, red, and near-infrared energy. As expected, the green light is reflected slightly more than the red light for the water pixel (3) while most of the incident near-infrared radiant flux is absorbed by the water, causing the brightness value to approach zero.

A second example demonstrates the information content of hyperspectral remote sensing data. The spectral profiles for features located in HyMap hyperspectral data of the Debordiu residential colony near North Inlet, SC, are shown in Figure 8-8. The hyperspectral dataset consists of 116 atmospherically corrected bands

of data at 3 × 3 m spatial resolution. Please note that the data have been calibrated to percent reflectance units. Only 3 of the 116 bands are displayed, including band 9 (green), band 15 (red), and band 40 (near-infrared). The spectral profile for the golf-putting green (1) exhibits all the characteristics of a well-calibrated hyperspectral vegetation pixel. There is chlorophyll absorption in the blue and red portions of the spectrum and significant reflectance throughout the near-infrared part of the spectrum. Reflectance in the middle-infrared bands in the region 1.55 – 1.75 μm and 2.08 – 2.35 μm is also strong. The atmospheric water absorption bands at 1.4 and 1.9 μm are evident in all the spectra. The spectral profile for sand (2) is high throughout the visible bands (blue, green, and red), causing it to appear bright white to human observers. The spectral profile for a residential rooftop (3) reflects relatively lower amounts of blue, green, and red energy throughout the visible spectrum, causing it to appear gray. Finally, water (4) absorbs more and more incident energy as we progress from the visible into the near-and middle-infrared portions of the spectrum.



Contrast Enhancement

Remote sensing systems record reflected and emitted radiant flux exiting from Earth's surface materials. Ideally, one material would reflect a tremendous amount of energy in a certain wavelength and another material would reflect much less energy in the same wavelength. This would result in *contrast* between the two types of material when recorded by the remote sensing system. Unfortunately, different materials often reflect similar amounts of radiant flux throughout the visible, near-infrared, and middle-infrared portions of the electromagnetic spectrum, resulting in relatively *low-contrast* imagery. In addition, besides this obvious low-contrast characteristic of biophysical materials, there are cultural factors at work. For example, people in developing countries often construct urban areas using natural building materials (e.g., wood, sand, silt, clay). This can cause urbanized areas in developing countries to have about the same reflectance characteristics as the neighboring countryside. Conversely, urban infrastructure in developed countries is usually composed of concrete, asphalt, and fertilized green vegetation. This typically causes urbanized areas in developed countries to have reflectance characteristics significantly different from the surrounding countryside.

An additional factor in the creation of low-contrast remotely sensed imagery is the sensitivity of the detectors. For example, the detectors on most remote sensing systems are designed to record a relatively wide range of scene brightness values (e.g., 0 to 255) without becoming saturated. Saturation occurs if the radiometric sensitivity of a detector is insufficient to record the full range of intensities of reflected or emitted energy emanating from the scene. The Landsat TM detectors, for example, must be sensitive to reflectance from diverse biophysical materials such as dark volcanic basalt outcrops or snow (possibly represented as *BV*s of 0 and 255, respectively). However, very few scenes are composed of brightness values that use the full sensitivity range of the Landsat TM detectors. Therefore, this results in relatively low-contrast imagery, with original brightness values that often range from approximately 0 to 100.

To improve the contrast of digital remotely sensed data, it is desirable to use the entire brightness range of the display or hard-copy output device (discussed in Chapter 5). Digital methods may be more useful than photographic techniques for contrast enhancement because of the precision and wide variety of processes that can be applied to the imagery. There are linear and nonlinear digital contrast-enhancement techniques.

Linear Contrast Enhancement

Contrast enhancement (also referred to as **contrast stretching**) expands the original input brightness values to make use of the total dynamic range or sensitivity of the output device. To illustrate the linear contrast-stretching process, consider the Charleston, SC, TM band 4 image produced by a sensor system whose image output levels can vary from 0 to 255. A histogram of this image is provided (Figure 8-9a). We will assume that the output device (a high-resolution display) can display 256 shades of gray (i.e., $quant_k = 255$). The histogram and associated statistics of this band 4 subimage reveal that the scene is composed of brightness values ranging from a minimum of 4 (i.e., $min_4 = 4$) to a maximum value of 105 (i.e., $max_4 = 105$), with a mean of 27.3 and a standard deviation of 15.76 (refer to Table 4-7). When these data are displayed on the screen without any contrast enhancement, we use less than one-half of the full dynamic range of brightness values that could be displayed (i.e., brightness values between 0 and 3 and between 106 and 255 are not used). The image is rather dark, low in contrast, with no distinctive bright areas (Figure 8-9a). It is difficult to visually interpret such an image. A more useful display can be produced if we expand the range of original brightness values to use the full dynamic range of the video display.

Minimum–Maximum Contrast Stretch

Linear contrast enhancement is best applied to remotely sensed images with Gaussian or near-Gaussian histograms, that is, when all the brightness values fall generally within a single, relatively narrow range of the

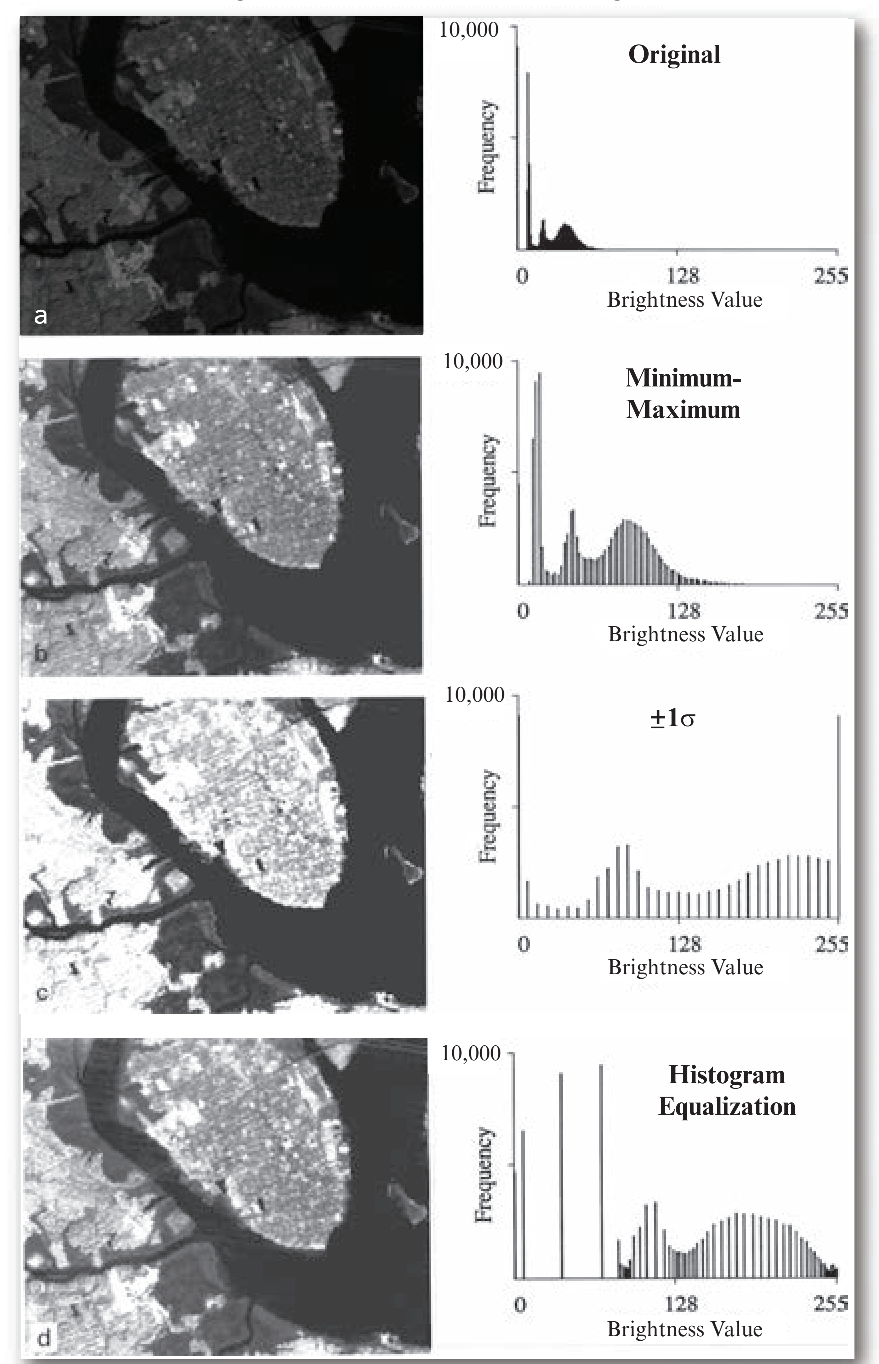


FIGURE 8–9 a) Original Landsat Thematic Mapper band 4 data of Charleston, SC, and its histogram. This image has *not* been contrast stretched. b) Minimum–maximum contrast stretch applied to the data and the resultant histogram. c) One standard deviation (– 1 σ) linear contrast stretch applied to the data and the resultant histogram. d) Application of histogram equalization and the resultant histogram. Original imagery courtesy of NASA.

histogram and only one mode is apparent. Unfortunately, this is rarely the case, especially for scenes that contain both land and water bodies. To perform a linear contrast enhancement, the analyst examines the image statistics and determines the minimum and maximum brightness values in band k, $\min_k$ and $\max_k$, respectively. The output brightness value, BV_{out}, is computed according to the equation:

$$BV_{out} = \left(\frac{BV_{in} - \min_k}{\max_k - \min_k}\right)\text{quant}_k, \qquad (8.1)$$

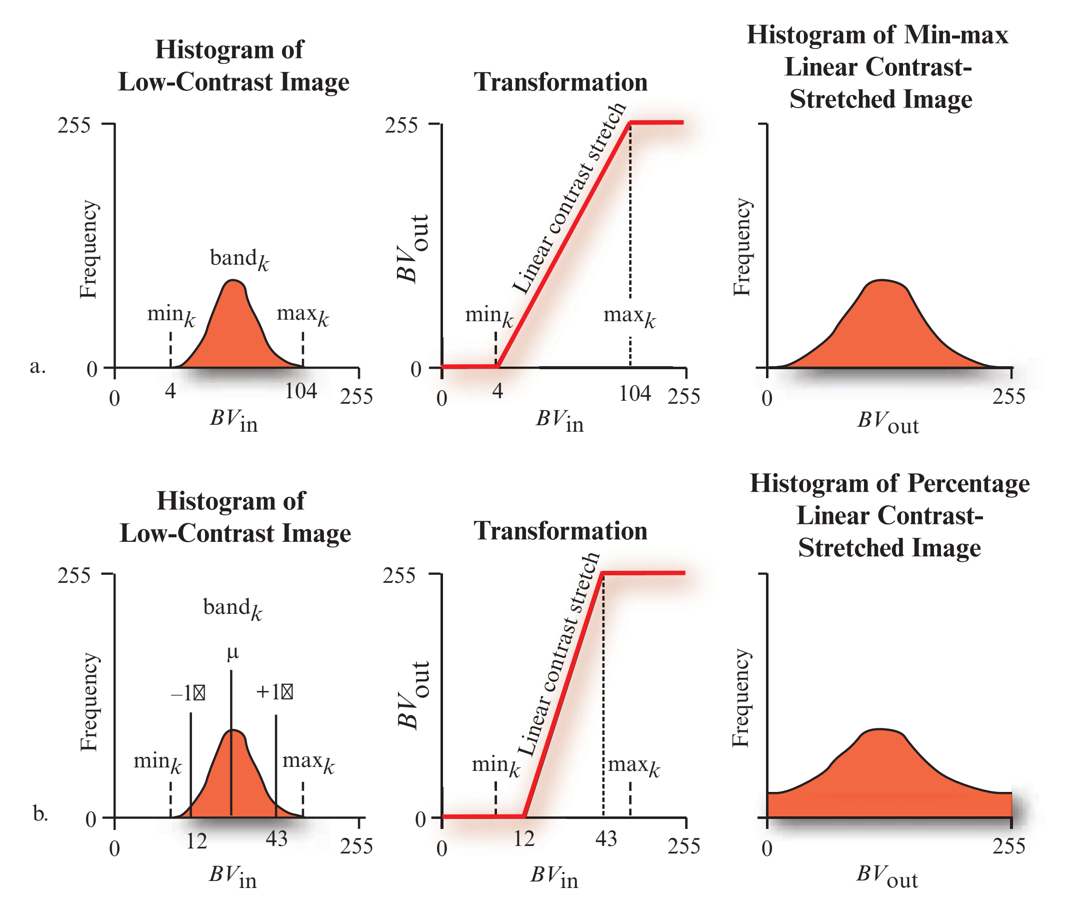


FIGURE 8–10 a) The result of applying a *minimum–maximum* contrast stretch to normally distributed remotely sensed data. The histograms before and after the transformation are shown. The minimum and maximum brightness values encountered in band *k* are $\min_k$ and $\max_k$, respectively. b) Theoretical result of applying a ±1 standard deviation *percentage linear* contrast stretch. This moves the $\min_k$ and $\max_k$ values ±34% from the mean into the tails of the distribution.

where BV_{in} is the original input brightness value and quant_k is the maximum value of the range of brightness values that can be displayed on the CRT (e.g., 255). In the Charleston, SC, example, any pixel with a BV_{in} of 4 would now have a BV_{out} of 0, and any pixel with a BV_{in} of 105 would have a BV_{out} of 255:

$$BV_{out} = \left(\frac{4-4}{105-4}\right)255$$

$$BV_{out} = 0$$

$$BV_{out} = \left(\frac{105-4}{105-4}\right)255$$

$$BV_{out} = 255\,.$$

The original brightness values between 5 and 104 would be linearly distributed between 0 and 255, respectively. The application of this enhancement to the Charleston TM band 4 data is shown in Figure 8-9b. This is commonly referred to as a *minimum–maximum contrast stretch*. Most image processing systems provide for the display of a before-and-after histogram, as well as a graph of the relationship between the input brightness value (BV_{in}) and the output brightness value (BV_{out}). For example, the histogram of the min–max contrast stretch discussed is shown in Figure 8-9b.

The logic of a *min-max linear contrast stretch* is shown diagrammatically in Figure 8-10a. Note the linear relationship between the brightness values of the input and output pixels and how the slope of the line would become increasingly steep as the minimum is increased or the maximum is decreased. The application of a minimum–maximum contrast stretch enhancement to the Savannah River predawn thermal infrared data is shown in Figure 8-11b.

Percentage Linear and Standard Deviation Contrast Stretching

Image analysts often specify $\min_k$ and $\max_k$ that lie a certain percentage of pixels from the mean of the histogram. This is called a *percentage linear contrast stretch.* If the percentage selected coincides with a standard deviation percentage, then it is called a *standard deviation*

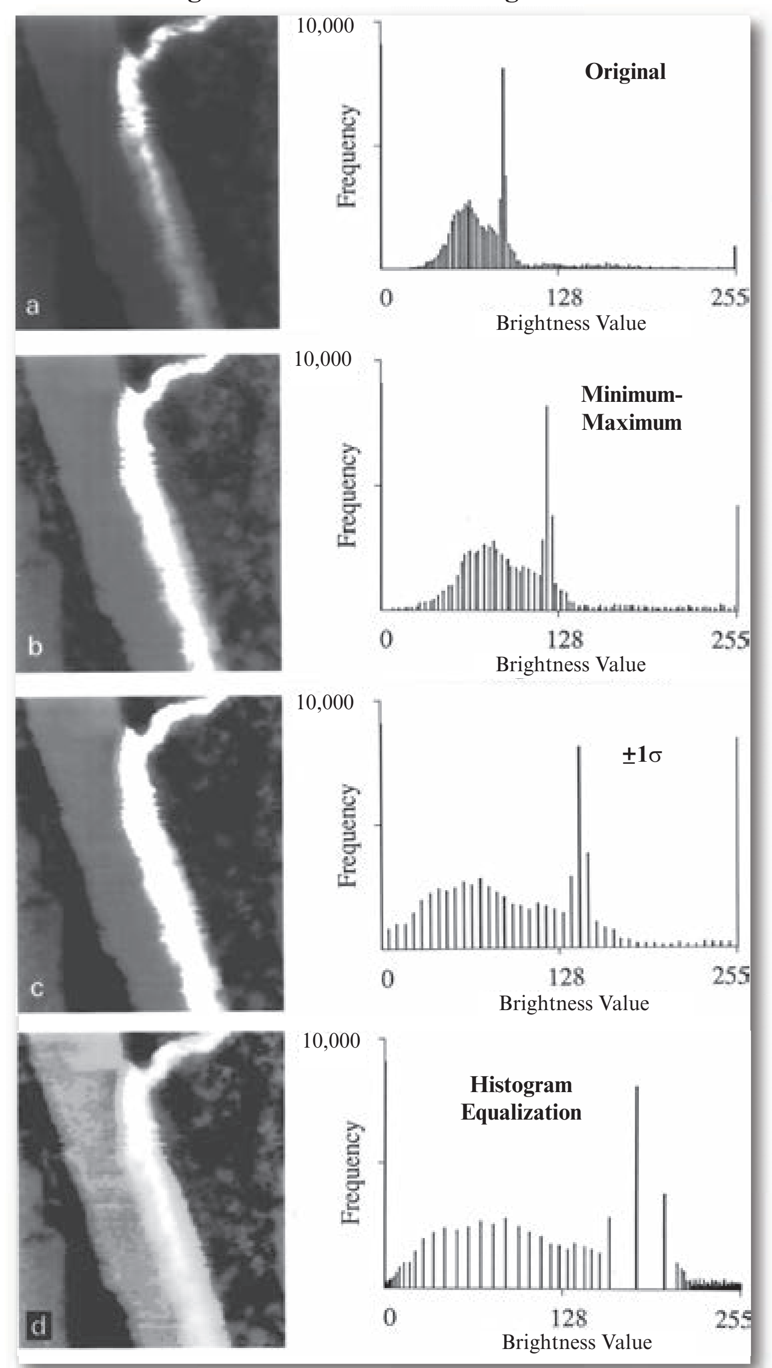


FIGURE 8–11 a) Original predawn thermal infrared data of the Savannah River, not contrast stretched, and its histogram. b) Minimum–maximum contrast stretch applied to the data and the resultant histogram. c) One standard deviation (–1σ) percentage linear contrast stretch applied to the data and resultant histogram. d) Application of histogram equalization and resultant histogram.

contrast stretch. For normal distributions, 68% of the observations lie within –1 standard deviation of the mean, 95.4% of all observations lie within –2 standard deviations, and 99.73% within –3 standard deviations. Consider applying a ±1 standard deviation contrast stretch to the Charleston, SC, Landsat TM band 4 data. This would result in $\min_k = 12$ and $\max_k = 43$. All values between 12 and 43 would be linearly contrast stretched to lie within the range 0 to 255. All values between 0 and 11 are now 0, and those between 44 and 255 are set to 255. This results in more pure black-and-white pixels in the Charleston, SC, scene, dramatically increasing the contrast of the image, as shown in Figure 8-9c. The information content of the pixels that

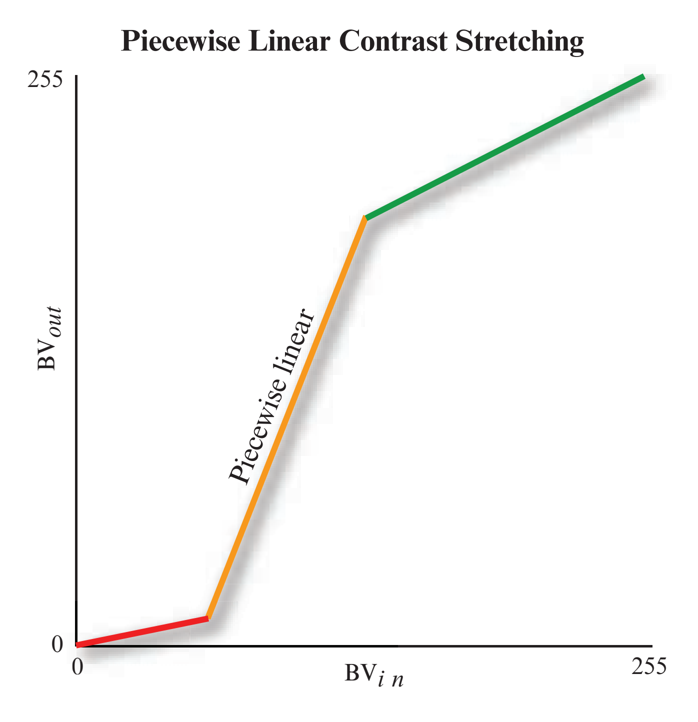


FIGURE 8–12 Logic of a piecewise linear contrast stretch for which three selective pieces of the histogram are linearly contrast stretched. Note that the slope of the linear contrast enhancement changes.

saturated at 0 and 255 is lost. The slope of a percentage linear or standard deviation contrast stretch is greater than for a simple min–max contrast stretch (refer to Figure 8-10b).

The results of applying a – 1 standard deviation linear contrast stretch to the thermal plume data are shown in Figure 8-11c along with the histogram. The – 1 standard deviation contrast stretch effectively "burns out" the thermal plume, yet provides more detail about the temperature characteristics of vegetation on each side of the river.

Piecewise Linear Contrast Stretch

When the histogram of an image is not Gaussian (i.e., it is bimodal, trimodal, etc.), it is possible to apply a piecewise linear contrast stretch to the imagery of the type shown in Figure 8-12. Here the analyst identifies a number of linear enhancement steps that expand the brightness ranges in the modes of the histogram. In effect, this corresponds to setting up a series of min_k and max_k and using Equation 8.1 within user-selected regions of the histogram. This powerful contrast enhancement method should be used when the analyst is intimately familiar with the various modes of the histogram and what they represent in the real world. Such contrast-stretched data are rarely used in subsequent image classification.

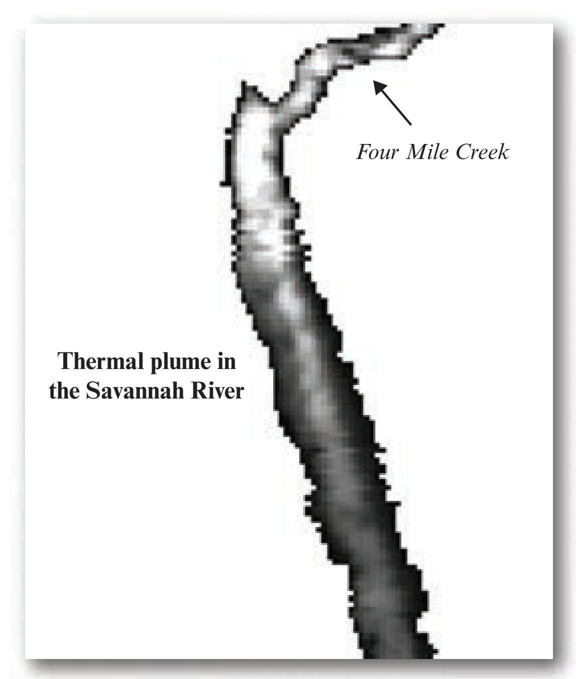
To perform piecewise linear contrast enhancement, the analyst normally views a) the raw image and b) a display of the histogram of the raw image superimposed with an input–output line that runs diagonally from the lower left to the upper right in the display. The analyst then interactively adjusts the length and slope of *n* mutually exclusive contrast stretches along the input–output line.

To illustrate the process, consider the two piecewise linear contrast enhancements found in Figure 8-13. The Savannah River thermal plume is composed primarily of values from 81 to 170. A special contrast stretch to highlight just the plume is demonstrated in Figures 8-13ab (note the red lines). Upland and ambient Savannah River water brightness values from 0 to 80 are sent to 255 (white), values from 81 to 170 are linearly contrast stretched to have values from 0 to 255, and all values from 171 to 255 now have a value of 255 (white). Conversely, if we wanted to contrast stretch the image to highlight the spectral characteristics of the ambient Savannah River, we might use the logic shown in Figures 8-13cd. In this example, the spike in the original histogram associated with just the ambient Savannah River water has been singled out and linearly contrast stretched to have values from 0 to 255. All the land and thermal plume pixels are set to a value of 0 and appear black.

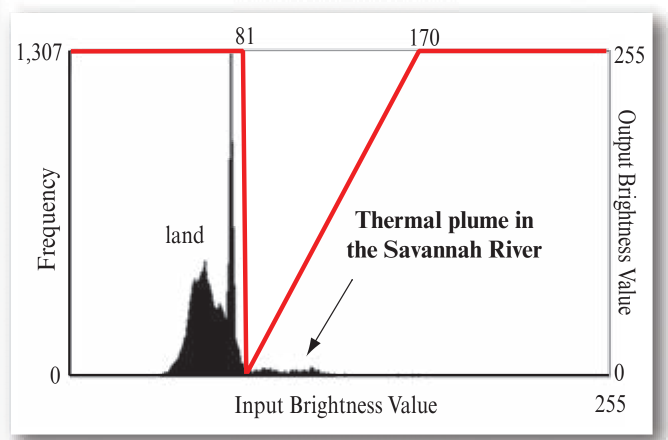
Nonlinear Contrast Enhancement

Nonlinear contrast enhancements may also be applied. One of the most useful enhancements is *histogram equalization*. The algorithm passes through the individual bands of the dataset and assigns approximately an equal number of pixels to each of the user-specified output grayscale classes (e.g., 32, 64, 256). Histogram equalization applies the greatest contrast enhancement to the most populated range of brightness values in the image. It automatically reduces the contrast in the very light or dark parts of the image associated with the tails of a normally distributed histogram.

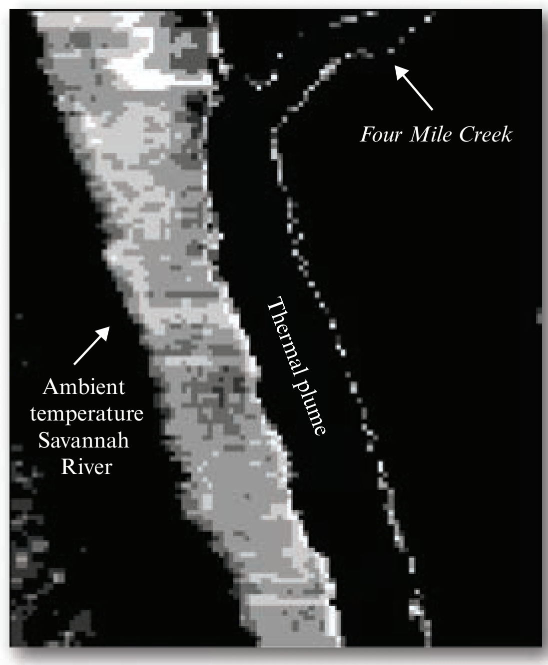
Histogram equalization is found in many image processing systems because it requires very little information from the analyst to implement (usually just the number of output brightness value classes desired and the bands to be equalized), yet it is often very effective. Because of its wide availability, it is instructive to review how the equalization takes place using a hypothetical dataset (Gonzalez and Wintz, 1977). For example, consider an image that is composed of 64 rows and 64 columns (4,096 pixels) with the range of brightness values that each pixel can assume, $quant_k$, limited to 0 through 7 (Table 8-2). A histogram of this hypothetical image is shown in Figure 8-14a and the frequency of occurrence of the individual brightness



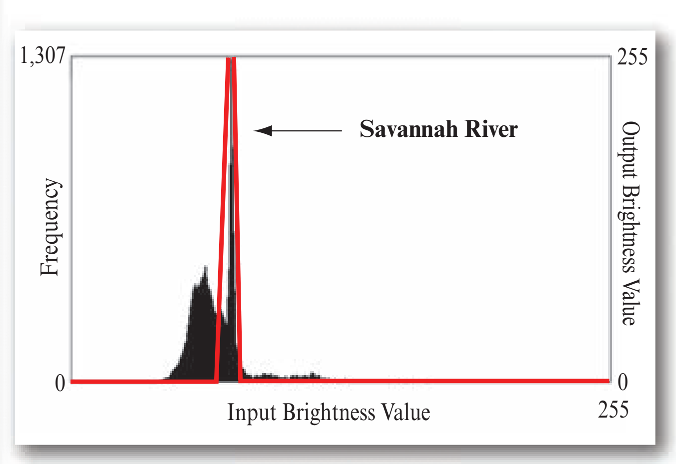
a. Thermal plume enhanced.



b. Enhancing the thermal plume in the Savannah River.



c. Savannah River enhanced.



d. Enhancing the Savannah River.

FIGURE 8–13 a,b) Thermal infrared imagery of the Savannah River enhanced using piecewise linear contrast stretching to highlight the thermal plume at the expense of the ambient river water and the surrounding landscape. c,d) Piecewise linear contrast stretching to enhance the Savannah River at the expense of the thermal plume and the surrounding landscape.

values, $f(BV_i)$, is summarized in Table 8-2. For example, there are 790 pixels in the scene with a brightness value of 0 (i.e., $f[BV_0] = 790$) and 1,023 pixels with a brightness value of 1 (i.e., $f[BV_1] = 1{,}023$). We can compute the probability of the ith brightness value, p_i, by dividing each of the frequencies, $f(BV_i)$, by the total number of pixels in the scene (i.e., $n = 4{,}096$). Thus, the probability of encountering a pixel with a brightness value of 0 in the scene is approximately 19% (i.e., $p_0 = f[BV_i]/n = 790/4{,}096 = 0.19$). A plot of the probability of occurrence of each of the eight brightness values for the hypothetical scene is shown in Figure 8-14b. This particular histogram has a large number of pixels with low brightness values (0 and 1), making it a relatively low-contrast scene.

TABLE 8–2 Statistics for a 64 ⊠ 64 hypothetical image with Brightness Values from 0 to 7 (n = 4,096 pixels) (modified from Gonzalez and Wintz, 1977).

Brightness Value, BV_i	L_i	Frequency $f(BV_i)$	Probability $p_i = f(BV_i)/n$
BV_0	0/7 = 0.00	790	0.19
BV_1	1/7 = 0.14	1023	0.25
BV_2	2/7 = 0.28	850	0.21
BV_3	3/7 = 0.42	656	0.16
BV_4	4/7 = 0.57	329	0.08
BV_5	5/7 = 0.71	245	0.06
BV_6	6/7 = 0.85	122	0.03
BV_7	7/7 = 1.00	81	0.02

The next step in the process is to compute a transformation function k_i for each brightness value. One way to conceptualize the histogram equalization process is to use the notation shown in Table 8-3. For each brightness value level BV_i in the quant$_k$ range of 0 to 7 of the original histogram, a new cumulative frequency value k_i is calculated:

$$k_i = \sum_{i=0}^{\text{quant}_k} \frac{f(BV_i)}{n} \qquad (8.2)$$

where the summation counts the frequency of pixels in the image with brightness values equal to or less than BV_i, and n is the total number of pixels in the entire scene (4,096 in this example). The histogram equalization process iteratively compares the transformation function k_i with the original values of L_i to determine which are closest in value. The closest match is reassigned to the appropriate brightness value. For example, in Table 8-3 we see that $k_0 = 0.19$ is closest to $L_1 =$ 0.14. Therefore, all pixels in BV_0 (790 of them) will be assigned to BV_1. Similarly, the 1,023 pixels in BV_1 will be assigned to BV_3, the 850 pixels in BV_2 will be assigned to BV_5, the 656 pixels in BV_3 will be assigned to BV_6, the 329 pixels in BV_4 will also be assigned to BV_6, and all 448 brightness values in $BV_{5–7}$ will be assigned to BV_7. The new image will have no pixels with brightness values of 0, 2, or 4. This is evident when evaluating the new histogram (Figure 8-14d). When analysts see such gaps in image histograms, it is usually a good indication that histogram equalization or some other operation has been applied.

Histogram-equalized versions of the Charleston TM band 4 data and the thermal plume data are found in Figure 8-9d and 8-11d, respectively. Histogram equalization is dramatically different from any other contrast enhancement because the data are redistributed according to the cumulative frequency histogram of the data, as described. Note that after histogram equalization, some pixels that originally had different values are now assigned the same value (perhaps a loss of information), while other values that were once very close together are now spread out, increasing the contrast between them. Therefore, this enhancement may improve the visibility of detail in an image, but it also alters the relationship between brightness values and image structure (Russ, 2011). For these reasons, it is not wise to extract texture or biophysical information from imagery that has been histogram equalized.

Another type of nonlinear contrast stretch involves scaling the input data *logarithmically,* as diagrammed in Figure 8-15. This enhancement has the greatest impact on the brightness values found in the darker part of the histogram. It could be reversed to enhance values in the brighter part of the histogram by scaling the input data using an inverse log function, as shown.

The selection of a contrast-enhancement algorithm depends on the nature of the original histogram and the elements of the scene that are of interest to the user. An experienced image analyst can usually identify an appropriate contrast-enhancement algorithm by examining the image histogram and then experimenting until satisfactory results are obtained. Most contrast enhancements cause some useful information to be lost. However, that which remains should be of value. Contrast enhancement is applied primarily to improve visual image analysis. It is *not* good practice to contrast stretch the original imagery and then use the enhanced imagery for computer-assisted classification, change detection, etc. Contrast stretching can distort the original pixel values, often in a nonlinear fashion.



Band Ratioing

Sometimes differences in brightness values from identical surface materials are caused by topographic slope and aspect, shadows, or seasonal changes in sunlight illumination angle and intensity. These conditions may hamper the ability of an interpreter or classification algorithm to identify correctly the surface materials or land use/land cover in a remotely sensed image. Fortunately, *ratio* transformations of the remotely sensed data can, in certain instances, be applied to reduce the effects of such environmental conditions. In addition

TABLE 8–3 Example of how a hypothetical 64 ✕ 64 image with brightness values from 0 to 7 is histogram equalized.

Frequency, $f(BV_i)$	790	1023	850	656	329	245	122	81
Original brightness value, BV_i	0	1	2	3	4	5	6	7
$L_i = \frac{\text{brightness value}}{n}$	0	0.14	0.28	0.42	0.57	0.71	0.85	1.0
Cumulative frequency transformation: $k_i = \sum_{i=0}^{\text{quant}_k} \frac{f(BV_i)}{n}$	$\frac{790}{4,096}$ =0.19	$\frac{1,813}{4,096}$ = 0.44	$\frac{2,663}{4,096}$ = 0.65	$\frac{3,319}{4,096}$ = 0.81	$\frac{3,648}{4,096}$ = 0.89	$\frac{3,893}{4,096}$ = 0.95	$\frac{4,015}{4,096}$ = 0.98	$\frac{4,096}{4,096}$ = 1.0
Assign original BV_i class to the new class it is closest to in value.	1	3	5	6	6	7	7	7

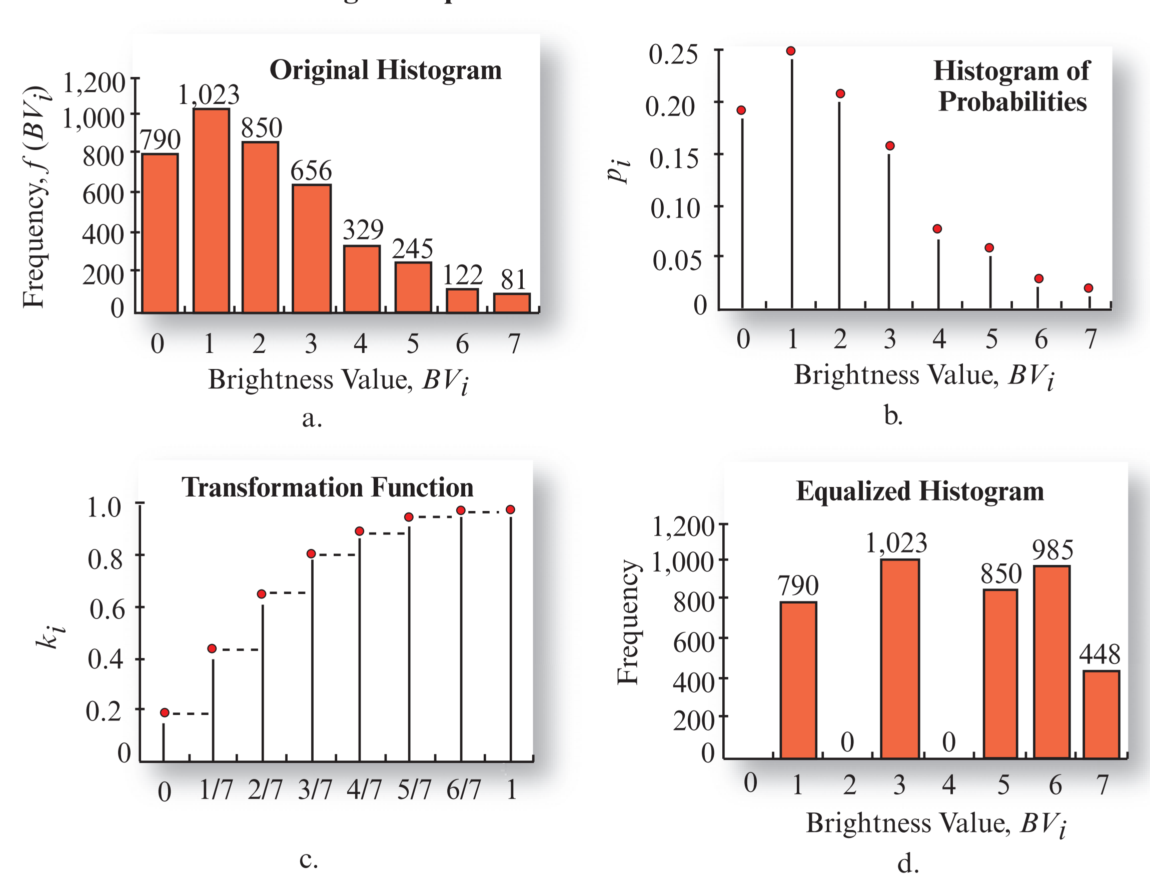


FIGURE 8–14 Histogram equalization process applied to hypothetical data (based on Gonzalez and Wintz, 1977). a) Original histogram showing the frequency of pixels in each brightness value. b) Original histogram expressed in probabilities. c) The transformation function. d) The equalized histogram showing the frequency of pixels in each brightness value.

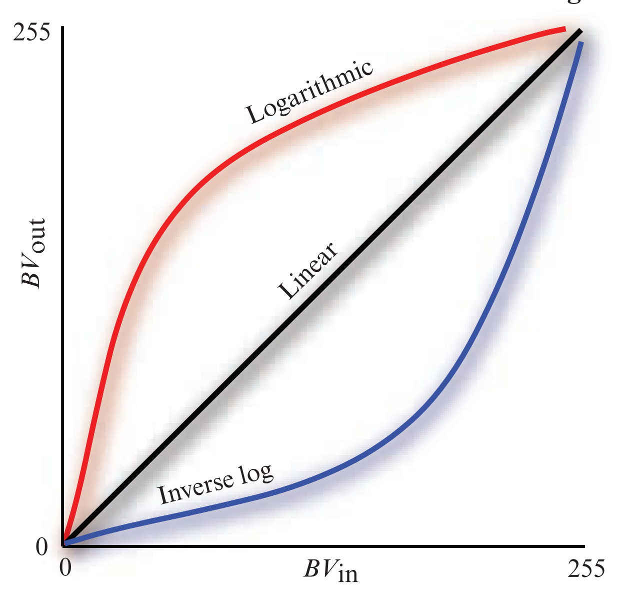


FIGURE 8–15 Logic of a linear, nonlinear logarithmic, and inverse logarithmic contrast stretch.

to minimizing the effects of environmental factors, ratios may also provide unique information not available in any single band that is useful for discriminating between soils and vegetation.

The mathematical expression of the ratio function is:

$$BV_{i,j,r} = \frac{BV_{i,j,k}}{BV_{i,j,l}} \tag{8.3}$$

where $BV_{i,j,r}$ is the output ratio value for the pixel at row *i*, column *j* and $BV_{i,j,k}$ and $BV_{i,j,l}$ are the brightness values at the same location in bands *k* and *l*, respectively. Unfortunately, the computation is not always simple because $BV_{i,j} = 0$ is possible. However, there are alternatives. For example, the mathematical domain of the function is 1/255 to 255 (i.e., the range of the ratio function includes all values beginning at 1/255, passing through 0, and ending at 255). The way to overcome this problem is simply to give any $BV_{i,j}$ with a value of 0 the value of 1. Alternatively, some like to add a small value (e.g., 0.1) to the denominator if it equals zero.

To represent the range of the function in a linear fashion and to encode the ratio values in a standard 8-bit format (values from 0 to 255), normalizing functions are applied. Using this normalizing function, the ratio value 1 is assigned the brightness value 128. Ratio values within the range 1/255 to 1 are assigned values between 1 and 128 by the function:

$$BV_{i,j,n} = \text{Int}\,[(BV_{i,j,r} \times 127) + 1]. \tag{8.4}$$

Ratio values from 1 to 255 are assigned values within the range 128 to 255 by the function:

$$BV_{i,j,n} = \text{Int}\left(128 + \frac{BV_{i,j,r}}{2}\right). \tag{8.5}$$

Deciding which two bands to ratio is not always a simple task. Often, the analyst simply displays various ratios and then selects the most visually appealing. The optimum index factor (OIF) and Sheffield Index (discussed in Chapter 5) can be used to identify optimum bands for band ratioing (Chavez et al., 1984; Sheffield, 1985). Crippen (1988) recommended that all data be atmospherically corrected and free from any sensor calibration problems (e.g., a detector is out of adjustment) before they are ratioed.

Numerous band ratios of Landsat Thematic Mapper data of Charleston, SC, obtained on February 3, 1994, are shown in Figure 8-16. A color-composite consisting of bands 4, 3, and 2 is shown in Figure 8-16a. Band ratios of the visible bands (1 through 3) are shown in Figures 8-16b,c, and e. Not surprisingly, because the visible bands are so highly correlated, it is difficult to distinguish between some of the major land covers such as wetland versus water. Band ratios of the near-infrared band 4 data with the visible bands are shown in Figure 8-16d, f, g, and h. The 3/4 band ratio is the inverse of the 4/3 band ratio. The 4/3 near-infrared/red ratio provides vegetation information that will be discussed in the section on vegetation indexes in this chapter. The brighter the pixel, the more vegetation or biomass present. Generally, the lower the correlation between the bands, the greater the information content of the band-ratioed image. For example, the ratio of bands 6 (thermal) and 4 (near-infrared) provides useful information (Figure 8-16k). Two color composites are displayed that incorporate three band-ratio images. For example, Figure 8-16o is a color composite of band ratios 1/4, 1/3, and 1/2 (RGB). Figure 8-16p is a color composite of band ratios 4/3, 1/3, and 6/4 (RGB). This band ratio is especially informative with water in light blue, wetland in light green, upland vegetation in yellow and red, and urbanized areas in shades of gray.

As you would expect, there are a tremendous number of band ratio images that can be derived from a multispectral and especially a hyperspectral image dataset. The goal is to create ratios based on the knowledge of some physical process or principle such as the relationship that exists for healthy vegetation, i.e., it always absorbs most of the incident red light while reflecting most of the incident near-infrared light. Thus, a near-infrared/red ratio is ideal for identifying the presence of vegetation in an image as demonstrated. Other more

Band Ratioing of Landsat Thematic Mapper Imagery of Charleston, SC, February 3, 1994

a. Original RGB = bands 4,3,2. b. band 1 / band 2. c. 1 / 3. d. 1 / 4.

e. 2 / 3. f. 4 / 2. g. 3 / 4. h. 4 / 3.

i. 4 / 5. j. 4 / 7. k. 6 / 4. l. 6 / 5.

m. 6 / 7. n. 7 / 4. o. RGB = 1/4, 1/3, 1/2. p. RGB = 4/3, 1/3, 6/4.

FIGURE 8–16 The ratio of various Landsat TM bands of Charleston, SC, obtained on February 3, 1994.

complex band ratio transformations are discussed in the vegetation index section later in this chapter.



Neighborhood Raster Operations

Local raster operations such as image reduction, magnification, band ratioing, etc. modify the values of each pixel in an image independent of the characteristics of neighboring pixels. Conversely, **neighborhood raster operations** modify the value of each focal pixel in the context of the values of the pixels surrounding it. Some of the most commonly used raster neighborhood types are shown in Figure 8-17. The *focal cell* (f) is usually at the center of the neighborhood being examined, as shown in Figure 8-17a.

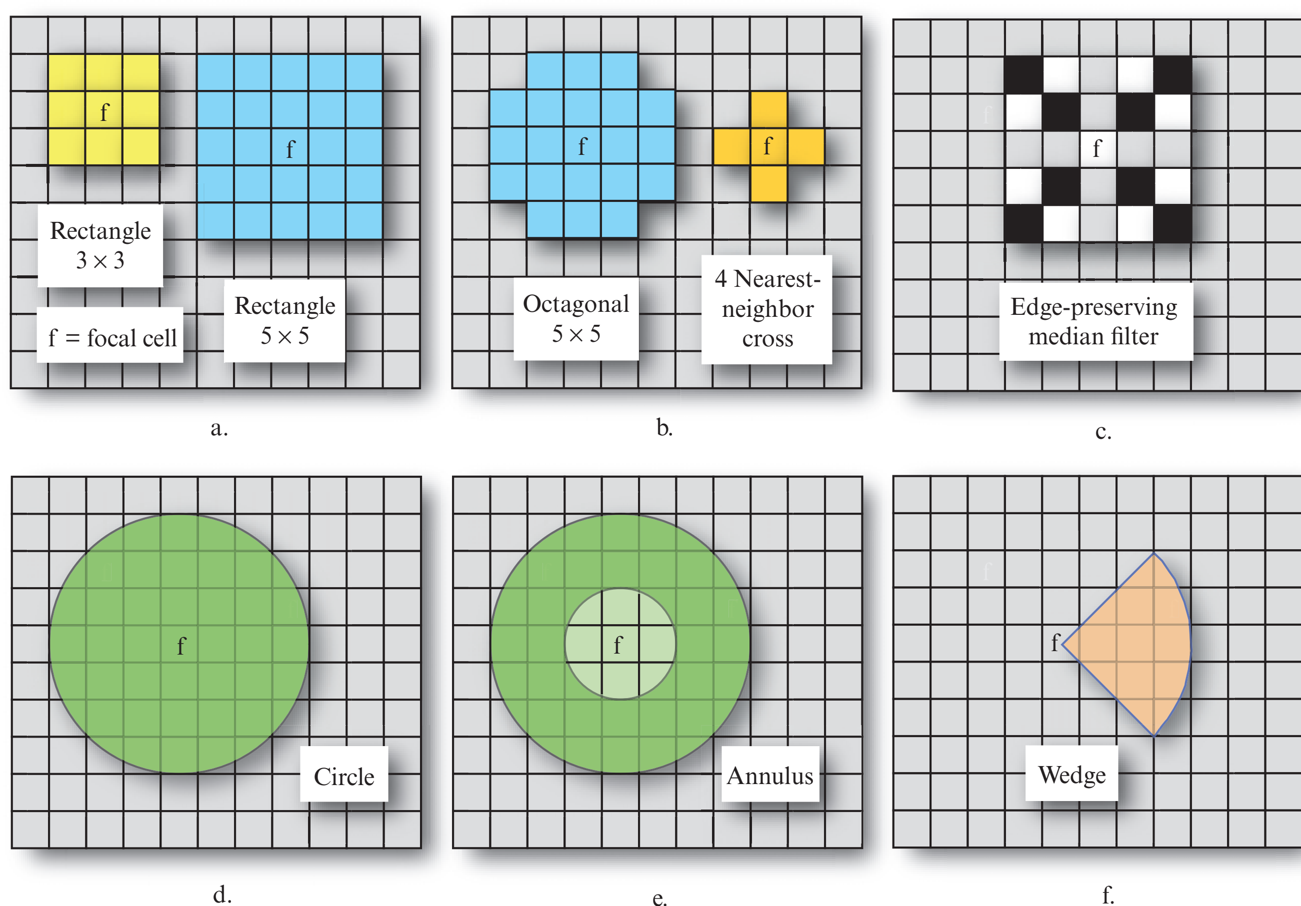


FIGURE 8–17 Examples of various convolution masks.

The raster rectangular neighborhood operators are probably the most heavily used, with 3 × 3 and 5 × 5 being the most prevalent (Figure 8-17a). The 5 × 5 octagonal, the 4 nearest-neighbor cross, and the edge-preserving median filter neighborhoods are also used when filtering images (Figure 8-17bc). Rectangular neighborhoods usually have an odd number of rows and columns so that there is always a central, focal pixel that is being evaluated. It is important to remember, however, that an image analyst can specify virtually any size of rectangular neighborhood, e.g., 9 × 9, 17 × 17, 25 × 25.

The logic of using a rectangular neighborhood to perform spatial filtering is shown in Figure 8-18. Note that in this example the 3 × 3 spatial moving window is stepped sequentially from left to right through the input raster dataset, line by line. Also note that the first row and column of the output matrix may not have any values when using a 3 × 3 neighborhood. In this case, the user must decide whether or not to leave the rows and columns blank or replicate the values from an adjacent row or column.

The circle neighborhood extends a user-specified radius away from the focal cell (Figure 8-17d). An annulus neighborhood centered on the focal cell includes the geographic area (i.e., pixels) bounded by the outer circle and the inner circle (Figure 8-17e). Pixels within the innermost circle are *not* included in the annulus computation. The wedge neighborhood encompasses a piece of a circle radiating out from the focal cell (Figure 8-17f).

Note that often only a part of a given pixel (cell) is included within the boundaries of a circular, annulus, or wedge neighborhood. When this occurs, the pixels are included if their centroids (i.e., the exact center of the pixels) lie within the neighborhood.

Qualitative Raster Neighborhood Modeling

Some of the most important uses of raster neighborhood analysis involves nominal-scale data such as land cover (e.g., Class A, B, and C) or ordinal-scaled data (e.g., Good =1, Adequate =2, and Poor = 3). However, unlike the previous section which only involved the examination of individual pixels in a single file or in multiple registered files, raster neighborhood analysis

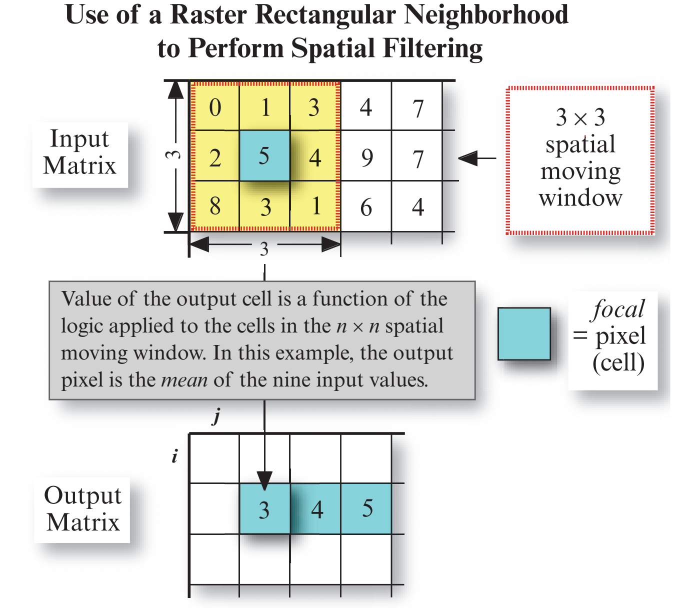


FIGURE 8–18 The logic of using a rectangular neighborhood to perform spatial filtering. Note that the spatial window is stepped sequentially from left to right through the input raster, line by line. Also note that the first row and column of the output matrix may not have any values when using a 3 × 3 neighborhood. In this case, the user must decide whether or not to leave them blank or replicate the values from an adjacent row or column.

usually examines a focal cell as well as a predetermined number of pixels surrounding the focal cell and uses this information to assign a new value to the focal cell in a new output file. A few of the qualitative measures that can be determined are shown in Figure 8-19a.

The application of a simple 3 × 3 window majority filter is shown in Figure 8-19a. Majority filters are often used to remove the salt-and-pepper noise associated with remote sensing-derived land cover maps. The result is a land cover thematic map that is much more pleasing to look at because an isolated pixel (e.g., water) that is totally surrounded by a majority of other types of land cover (e.g., forest) is assigned to the majority land cover class that surrounds it. Sometimes it is good to determine the majority or minority of digital number values within a spatial moving window (Figure 8-19a).

Perhaps it is important to determine the diversity of values found within the spatial moving window (Figure 8-19a). For example, a high diversity value for a pixel might indicate the presence of a large number of land cover types in a relatively small region (i.e., within the 3 × 3 window). Such diversity might be especially important for certain animals that require a diverse (heterogeneous) land cover habitat for survival. Conversely, some animals might thrive best in homogeneous (nondiverse) land cover (e.g., dense monoculture forest stands).

Quantitative Raster Neighborhood Modeling

Raster neighborhood analysis is often used to analyze interval and ratio-scaled data. Several of the most common quantitative univariate descriptive statistic measurements extracted from $n \times n$ spatial moving windows include the minimum, maximum, mean, and standard deviation as shown in Figure 8-19b.



Spatial Filtering

A characteristic of remotely sensed images is a parameter called **spatial frequency**, defined as the number of changes in brightness value per unit distance for any particular part of an image. If there are very few changes in brightness value over a given area in an image, this is commonly referred to as a low-frequency area. Conversely, if the brightness values change dramatically over short distances, this is an area of high-frequency detail. Because spatial frequency by its very nature describes the brightness values over a spatial *region*, it is necessary to adopt a spatial approach to extracting quantitative spatial information. This is done by looking at the local (neighboring) pixel brightness values rather than just an independent pixel value. This perspective allows the analyst to extract useful spatial frequency information from the imagery.

Spatial frequency in remotely sensed imagery may be enhanced or subdued using two different approaches. The first is *spatial convolution filtering* based primarily on the use of convolution masks. The procedure is relatively easy to understand and can be used to enhance low- and high-frequency detail, as well as edges in the imagery (Lo and Yeung, 2007). Another technique is *Fourier analysis,* which mathematically separates an image into its spatial frequency components. It is then possible interactively to emphasize certain groups (or bands) of frequencies relative to others and recombine the spatial frequencies to produce an enhanced image. We first introduce the technique of spatial convolution filtering and then proceed to the more mathematically challenging Fourier analysis.

Spatial Convolution Filtering

A **linear spatial filter** is a filter for which the value ($V_{i,j}$) at location i, j in the output image or map is a function of some weighted average (linear combination) of the

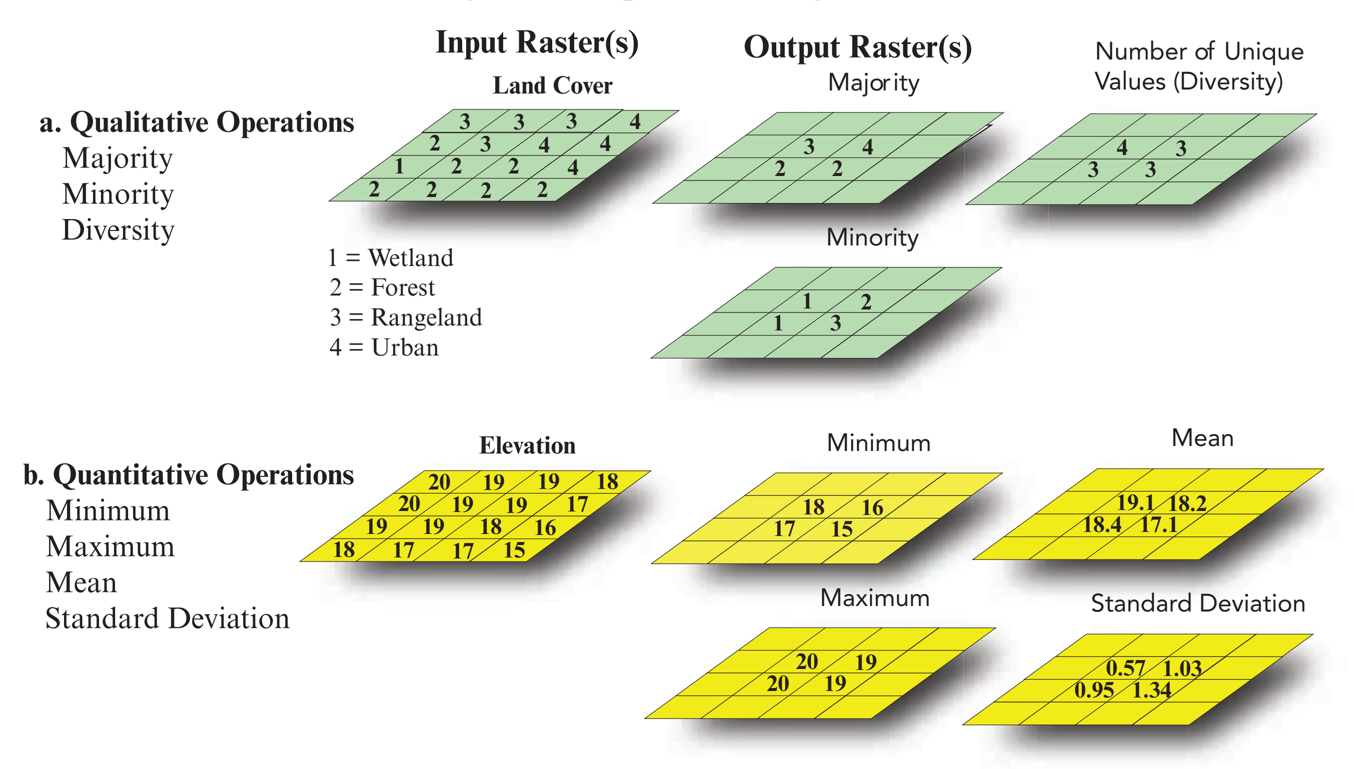


FIGURE 8–19 Neighborhood operations applied to a single raster dataset using a 3 × 3 spatial moving window. a) Simple qualitative operations. b) Simple quantitative operations.

values located in a particular spatial pattern around the *i, j* location in the input image or map (Figure 8-18 and 8-19). This process of evaluating the weighted neighboring pixel values is called *two-dimensional convolution filtering* (Pratt, 2014).

Sophisticated geographic information systems and remote sensing digital image processing software (e.g., ERDAS IMAGINE, ENVI) provide simple user interfaces that allow the analyst to specify the size of the convolution kernel (e.g., 3 × 3, 5 × 5) and the coefficients to be placed in the convolution kernel. Examples of ERDAS IMAGINE and ArcGIS ArcMap convolution filtering user interfaces are shown in Figure 8-20.

Spatial convolution filtering can be used to enhance low-frequency detail, high-frequency detail, and edges in raster digital images. The following discussion of spatial convolution filtering will be based on the application of various filters to a high-spatial-resolution color aerial photograph. Please remember, however, that the filters can also be applied to continuous raster thematic map data such as elevation, temperature, humidity, population density, etc. (Warner et al., 2009).

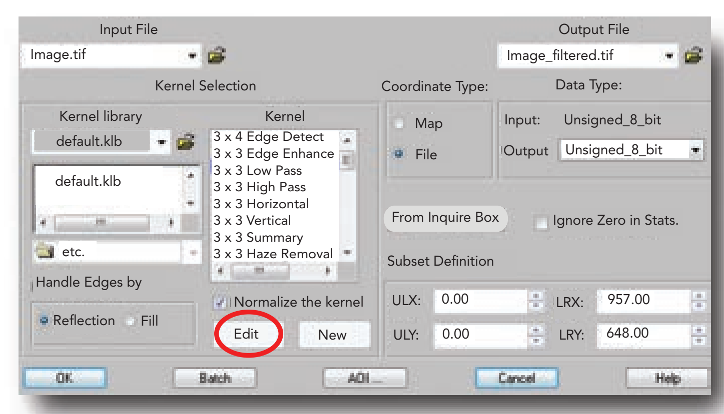
Low-frequency Filtering in the Spatial Domain

Raster enhancements that block or minimize the high-spatial frequency detail are called **low-frequency** or **low-pass** filters. The simplest low-frequency filter evaluates a particular input value, V_{in}, and the pixels surrounding the input pixel, and outputs a new value, V_{out}, that is the mean of this convolution. The size of the neighborhood convolution mask or kernel (n) is often 3 × 3, 5 × 5, 7 × 7, or 9 × 9. Examples of symmetric 3 × 3 and 5 × 5 convolution masks are shown in Figure 8-17a. The following discussion will focus primarily on the use of 3 × 3 convolution masks with nine coefficients, c_i, defined at the following locations:

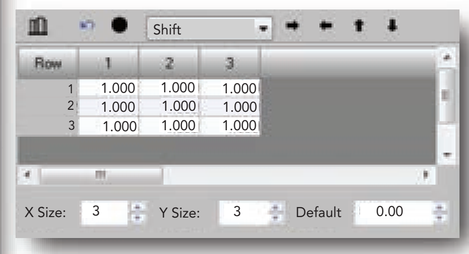
$$\text{Convolution mask template} = \begin{bmatrix} c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 \\ c_7 & c_8 & c_9 \end{bmatrix} \quad (8.6)$$

For example, the coefficients in a low-frequency convolution mask are usually set equal to 1 (e.g., Figure 8-20ab and Table 8-4):

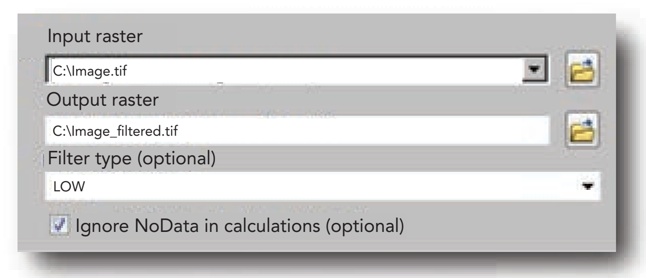
Spatial Convolution Filtering User Interfaces



a. ERDAS IMAGINE convolution interface.



b. ERDAS IMAGINE 3 × 3 low-pass filter.



c. ArcGIS ArcMap filter interface.

FIGURE 8–20 a) An example of the ERDAS IMAGINE convolution user interface. b) The analyst can use the edit menu to input the desired coefficients in an $n \times n$ kernel. In this case, a 3 × 3 low-pass filter is being prepared (interfaces courtesy of Hexagon Geospatial). c) Spatial filtering using the ArcGIS ArcMap spatial analyst filtering interface. A low-pass filter has been selected (interface courtesy of Esri, Inc.).

$$\text{Low-frequency filter} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{8.7}$$

The coefficients, c_i, in the mask template are multiplied by the following individual values (V_i) in the input digital image:

$$\text{Mask template} = \begin{matrix} c_1 \times V_1 & c_2 \times V_2 & c_3 \times V_3 \\ c_4 \times V_4 & c_5 \times V_5 & c_6 \times V_6 \\ c_7 \times V_7 & c_8 \times V_8 & c_9 \times V_9 \end{matrix} \tag{8.8}$$

The primary input pixel under investigation at any one time is $V_5 = V_{i,j}$. Convolution of the low-frequency filter (with all coefficients equal to 1) and the original image data will result in a low-frequency filtered image where:

$$LFF_{5,\text{out}} = \text{Int}\frac{\sum_{i=1}^{n=9} c_i \times V_i}{n} = \text{Int}\left(\frac{V_1 + V_2 + V_3 + \dots + V_9}{9}\right) \tag{8.9}$$

The spatial moving average then shifts to the next pixel, where the average of all nine digital number values is computed. This operation is repeated for every pixel in the input image (Figure 8-18). Image smoothing is useful for removing “salt and pepper” noise in raster data. This simple smoothing operation will, however, blur the image, especially at the edges of objects. The blurring becomes more severe as the size of the kernel increases.

A high spatial resolution (6 × 6 in.) normal color digital aerial photograph of a residential area in Germany is shown in Figure 8-21a. It was obtained by an unmanned aerial vehicle (UAV) developed by senseFly, Inc. Application of a **low-frequency filter** (Equation 8.7 and Table 8-4) to the red band of the residential image is shown in Figure 8-21b. Note how the image becomes blurred, suppressing the high-frequency detail. Only the general trends are allowed to pass through the low-pass filter. In a heterogeneous, high-frequency urban environment, a high-frequency filter usually provides superior results.

Three spatial filtering algorithms (median, minimum/maximum, and Olympic) do not use coefficients within the $n \times n$ neighborhood being examined. A **median filter** has certain advantages when compared with weighted convolution filters (Russ, 2011), including 1) it does not shift boundaries, and 2) the minimal degradation to edges allows the median filter to be applied

TABLE 8–4 Selected low- and high-frequency filters and linear and nonlinear edge enhancement filters with their convolution mask coefficients.

Convolution mask template = $\begin{bmatrix} c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 \\ c_7 & c_8 & c_9 \end{bmatrix}$ Example: Low-frequency filter = $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	**Example**
Spatial Filtering										
Low-frequency (LFF)	1	1	1	1	1	1	1	1	1	Figure 8-21b
High-frequency (HFF)	1	–2	1	–2	5	–2	1	–2	1	Figure 8-21d
Linear Edge Enhancement										
Emboss Northwest	0	0	1	0	0	0	–1	0	0	Figure 8-22a
Emboss East	0	0	0	1	0	–1	0	0	0	---
ArcGIS Edge Enhancement	–0.7	–1	–0.7	–1	6.8	–1	–0.7	–1	–0.7	---
Compass North	1	1	1	1	–2	1	–1	–1	–1	---
Compass NE	1	1	1	–1	–2	1	–1	–1	1	Figure 8-22b
Compass East	–1	1	1	–1	–2	1	–1	1	1	---
Compass SE	–1	–1	1	–1	–2	1	1	1	1	---
Compass South	–1	–1	–1	1	–2	1	1	1	1	
Compass SW	1	–1	–1	1	–2	–1	1	1	1	
Compass West	1	1	–1	1	–2	–1	1	1	–1	
Compass NW	1	1	1	1	–2	–1	1	––1	–1	
Vertical Edges	–1	0	1	–1	0	1	–1	0	1	
Horizontal Edges	–1	–1	–1	0	0	0	1	1	1	
Diagonal Edges	0	1	1	–1	0	1	–1	–1	0	
Nonlinear Edge Enhancement										
Laplacian 4	0	–1	0	–1	4	–1	0	–1	0	Figure 8-22c
Laplacian 5	0	–1	0	–1	5	–1	0	–1	0	Figure 8-22d
Laplacian 7	1	1	1	1	–7	1	1	1	1	
Laplacian 8	–1	–1	–1	–1	8	–1	–1	–1	–1	

Spatial Filtering of Raster Data





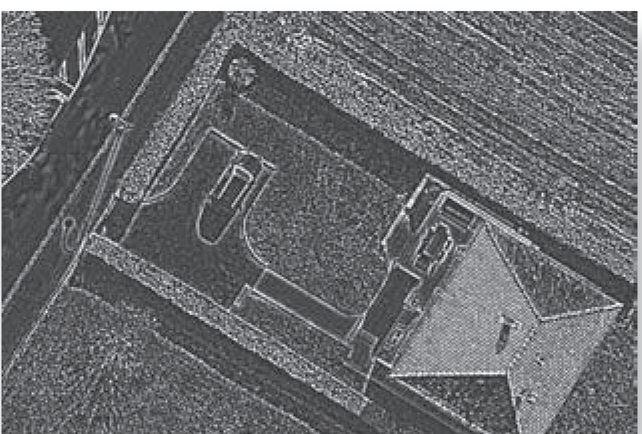
a. Original contrast stretched.



b. Low-frequency filter applied to the red band.



c. Median filter.



d. High-frequency sharp edge filter applied to the red band.

FIGURE 8–21 a) High spatial resolution color digital aerial photography of a residential area in Germany (original digital aerial photography courtesy of Sensefly, LLC). b) Low-frequency filter (LFF) applied to the red band. c) Median filter applied to the red, green, and blue bands. d) High-frequency sharp edge filter applied to the red band.

repeatedly, which allows fine detail to be erased and large regions to take on the same brightness value (often called *posterization*) (Figure 8-21c). The standard median filter will erase some lines in the image that are narrower than the half-width of the neighborhood and round or clip corners (Eliason and McEwen, 1990). An *edge-preserving median filter* (Nieminen et al., 1987) may be applied using the logic shown in Figure 8-17c, where 1) the median value of the black pixels is computed in a 5 × 5 array, 2) the median value of the gray pixels is computed, 3) these two values and the central original brightness value are ranked in ascending order, and 4) a final median value is selected to replace the central pixel. This filter generally preserves edges and corners.

Minimum or **maximum filters** examine the digital number values of adjacent pixels in a user-specified region (e.g., 3 × 3 pixels) and then replace the value of the center pixel with the minimum or maximum value encountered.

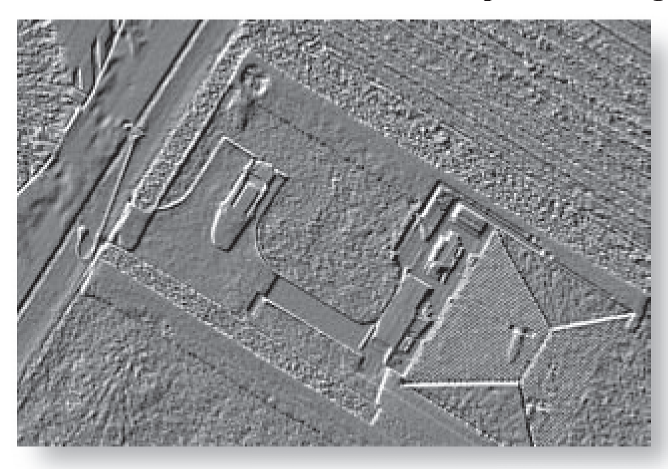
The **Olympic filter** is named after the system of scoring in Olympic events. Instead of using all nine elements in a 3 × 3 matrix, the highest and lowest values are dropped and the remaining values are averaged.

High-frequency Filtering in the Spatial Domain

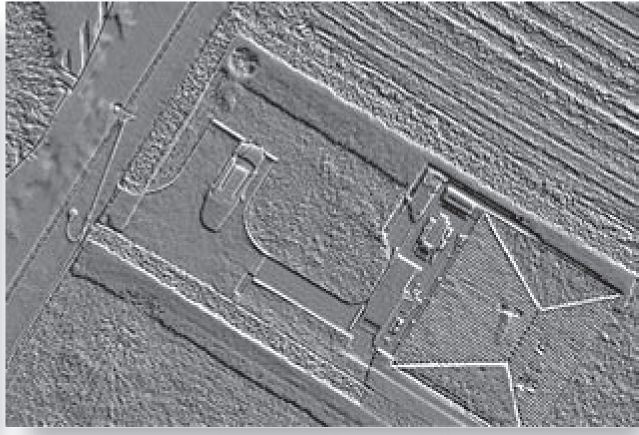
High-pass filtering is applied to remotely-sensed data to remove the slowly varying components and enhance the high-frequency local variations. One useful high-frequency filter ($HFF_{5,out}$) is computed by subtracting the output of the low-frequency filter ($LFF_{5,out}$) from twice the value of the original central pixel value, BV_5:

$$HFF_{5,out} = (2 \times BV_5) - LFF_{5,out}. \qquad (8.10)$$

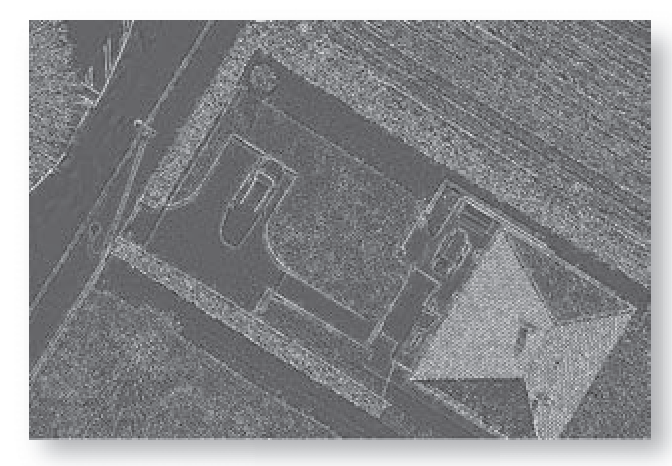
Spatial Filtering of Raster Data



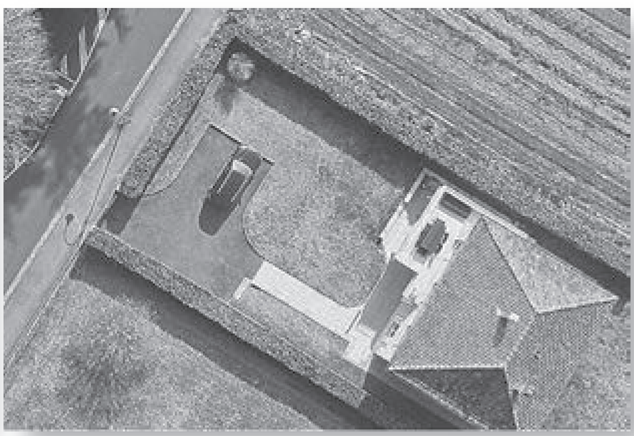
a. Emboss Northwest filter.



b. Compass Northeast filter.



c. Laplacian 4 filter.



d. Laplacian 5 filter.

FIGURE 8–22 Application of various convolution masks and logic to the red band of high spatial resolution imagery of a residential area in Germany. a) Emboss Northwest filter. b) Compass Northeast directional filter. c) Laplacian 4 edge enhancement contains information about the edges. d) Laplacian 5 edge enhancement adds the edge data back onto the original image (based on original digital aerial photography from Sensefly, LLC).

Brightness values tend to be highly correlated in a nine-element window. Thus, the high-frequency filtered image will have a relatively narrow intensity histogram. This suggests that the output from high-frequency filtered images often must be contrast stretched prior to visual analysis. High-frequency filtering is applied to raster maps or images to remove the slowly varying components and enhance the high-frequency local variations. One type of **high-frequency filter** that accentuates or sharpens edges uses the coefficients shown below and summarized in Table 8-4.

$$\text{High-frequency filter} = \begin{matrix} 1 & -2 & 1 \\ -2 & 5 & -2 \\ 1 & -2 & 1 \end{matrix} \tag{8.11}$$

The application of this high-frequency sharp edge filter to the residential aerial photograph red band is shown in Figure 8-21d.

Edge Enhancement in the Spatial Domain

Often, the most valuable information that may be derived from an image or map is contained in the edges surrounding objects of interest. **Edge enhancement** delineates these edges in an image or map and makes them more conspicuous and often easier to understand. Generally, what the eyes see as pictorial edges are simply sharp changes in digital number value between two adjacent pixels. The edges may be enhanced using either linear or nonlinear edge enhancement techniques.

Linear Edge Enhancement Edge enhancement is often performed by convolving the original data with a

weighted mask or kernel, as previously discussed. One of the most useful edge enhancements causes the edges to appear in a plastic shaded-relief format. This is often referred to as **embossing**. Embossed edges may be obtained by using embossing filters such as the Emboss East and Emboss Northwest shown below and in Table 8-4:

$$\text{Emboss East} = \begin{matrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{matrix} \tag{8.12}$$

$$\text{Emboss NW} = \begin{matrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{matrix} \tag{8.13}$$

The direction of the embossing is controlled by changing the value of the coefficients around the periphery of the mask. The plastic shaded-relief impression is pleasing to the human eye if shadows are made to fall toward the viewer. The red band of the residential image processed using the Emboss Northwest filter is shown in Figure 8-22a.

The ArcGIS high-pass edge enhancement coefficients are provided in Table 8-4. This particular filter removes low-frequency variation and highlights the boundary (edges) between different regions.

Compass gradient masks may be used to perform two-dimensional, discrete differentiation directional edge enhancement (Pratt, 2014). Table 8-4 lists the coefficients used in eight commonly used compass gradient masks. The compass names suggest the slope direction of maximum response. For example, the east gradient mask produces a maximum output for horizontal digital number value changes from west to east. The gradient masks have zero weighting (i.e., the sum of the mask coefficients is zero) (Pratt, 2014). This results in no output response over regions with constant brightness values (i.e., no edges are present). A Compass Northeast gradient mask applied to the residential image is shown in Figure 8-22b.

Richards (2013) identified four additional 3 × 3 filters that may be used to detect edges in images (vertical, horizontal, and diagonal). These filter coefficients are listed in Table 8-4.

Laplacian filters may be applied to imagery or continuous surface maps to perform edge enhancement. The Laplacian filter is a second derivative (as opposed to the gradient, which is a first derivative) and is invariant to rotation, meaning that it is insensitive to the direction in which the discontinuities (e.g., edges) run. The coefficients associated with four important 3 × 3 Laplacian filters are shown below and listed in Table 8-4 (Jahne, 2005; Pratt, 2014):

$$\text{Laplacian 4} = \begin{matrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{matrix} \tag{8.14}$$

$$\text{Laplacian 5} = \begin{matrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{matrix} \tag{8.15}$$

$$\text{Laplacian 7} = \begin{matrix} 1 & 1 & 1 \\ 1 & -7 & 1 \\ 1 & 1 & 1 \end{matrix} \tag{8.16}$$

$$\text{Laplacian 8} = \begin{matrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{matrix} \tag{8.17}$$

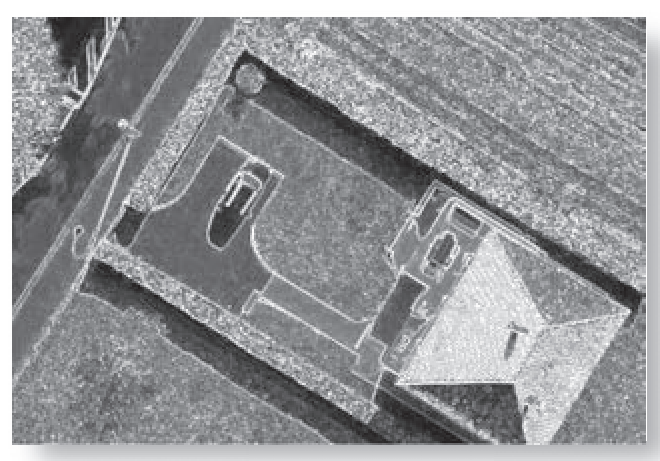
Sometimes by itself, the Laplacian image may be difficult to interpret. For example, consider the Laplacian 4 filter applied to the residential scene shown in Figure 8-22c. Therefore, some analysts prefer using a Laplacian edge enhancement that adds the edge information back onto the original map or image using the Laplacian 5 algorithm. The result of applying this enhancement to the residential scene is shown is Figure 8-22d.

The Laplacian 7 filter may be used to subtract the Laplacian edges from the original image, if desired. Subtracting the Laplacian edge enhancement from the original image restores the overall grayscale variation, which the human viewer can comfortably interpret. It also sharpens the image by locally increasing the contrast at discontinuities (Russ, 2011). The *Laplacian* operator generally highlights points, lines, and edges in the image and suppresses uniform and smoothly varying regions. Human vision physiological research suggests that we see objects in much the same way. Hence, the use of this operation has a more natural look than many of the other edge-enhanced images.

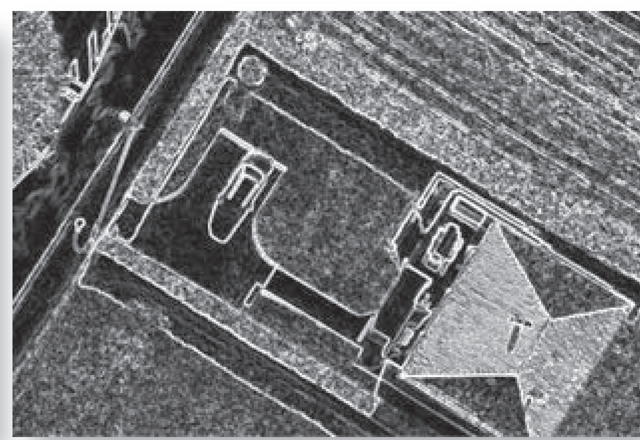
Nonlinear Edge Enhancement Nonlinear edge enhancements are performed using nonlinear combinations of pixels. For example, the **Sobel edge detector** is based on the 3 × 3 window numbering scheme previously described and is computed using the following equation:

$$\text{Sobel}_{5,\text{out}} = \sqrt{X^2 + Y^2} \tag{8.18}$$

Spatial Filtering of Raster Data



a. Sobel edge enhancement.



b. Roberts edge enhancement.

FIGURE 8–23 Application of two nonlinear edge enhancements to the red band of the high spatial resolution residential imagery. a) A Sobel edge enhancement. b) A Roberts edge enhancement (based on original digital aerial photography from Sensefly, LLC).

where

$$X = (V_3 + 2V_6 + V_9) - (V_1 + 2V_4 + V_7)$$

and

$$Y = (V_1 + 2V_2 + V_3) - (V_7 + 2V_8 + V_9)$$

This procedure detects horizontal, vertical, and diagonal edges. A Sobel edge enhancement of the residential scene is found in Figure 8-23a. The Sobel operator may also be computed by simultaneously applying the following 3 × 3 templates across the image (Jain, 1989):

$$X = \begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix} \qquad Y = \begin{matrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{matrix}$$

The **Roberts edge detector** is based on the use of only four elements of a 3 × 3 mask. The new pixel value at pixel location $V_{5,out}$ (refer to the 3 × 3 numbering scheme in Equation 8.6) is computed according to the equation:

$$\text{Roberts}_{5,\text{ out}} = X + Y \tag{8.19}$$

where

$$X = |V_5 - V_9|$$
$$Y = |V_6 - V_8|$$

A Roberts edge filter is applied to the residential scene in Figure 8-23b. The Roberts' operator also may be computed by simultaneously applying the following templates across the image (Jain, 1989):

$$X = \begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{matrix} \qquad Y = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{matrix}$$

The **Kirsch nonlinear edge enhancement** calculates the gradient at pixel location $BV_{i,j}$. To apply this operator, however, it is first necessary to designate a different 3 × 3 window numbering scheme than used in previous discussions:

Window numbering for Kirsch =

$$\begin{matrix} BV_0 & BV_1 & BV_2 \\ BV_7 & BV_{i,j} & BV_3 \\ BV_6 & BV_5 & BV_4 \end{matrix}$$

The algorithm applied is (Gil et al., 1983):

$$BV_{i,j} = \max\left\{1, \max_{i=0}^{7} [\text{Abs}(5S_i - 3T_i)]\right\} \tag{8.20}$$

where

$$S_i = BV_i + BV_{i+1} + BV_{i+2} \tag{8.21}$$

and

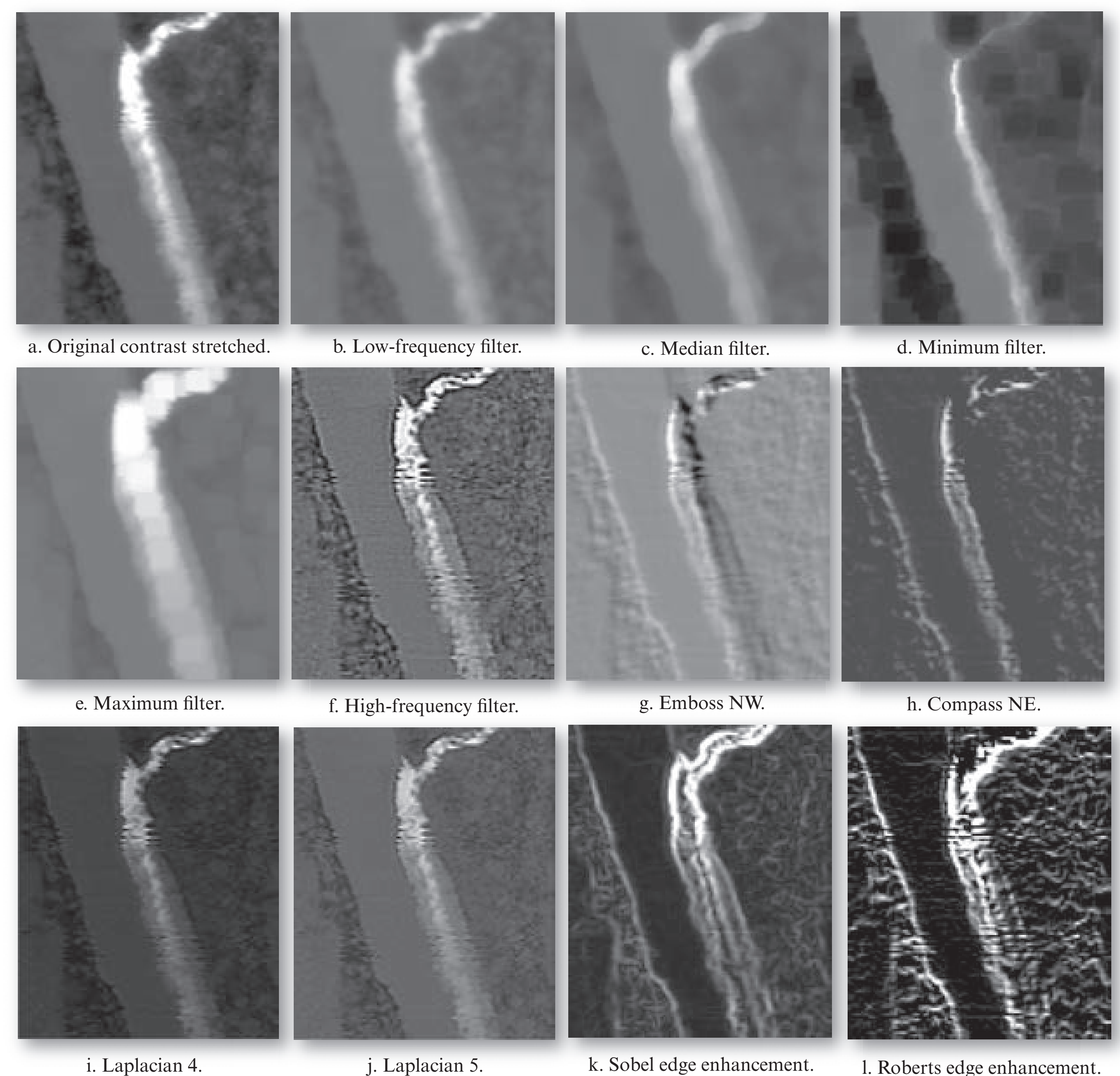


FIGURE 8–24 Application of various convolution masks and logic to the predawn thermal-infrared data to enhance low- and high-frequency detail: a) contrast stretched original image, b) low-frequency filter, c) median filter, d) minimum filter, e) maximum filter, f) high-frequency filter, g) emboss NW filter, h) compass NE filter, i) Laplacian 4 filter, j) Laplacian 5 filter, k) Sobel edge enhancement, and l) Robert's edge enhancement.

$$T_i = BV_{i+3} + BV_{i+4} + BV_{i+5} + BV_{i+6} + BV_{i+7}. \quad (8.22)$$

The subscripts of BV are evaluated modulo 8, meaning that the computation moves around the perimeter of the mask in eight steps. The edge enhancement computes the maximal compass gradient magnitude about input image point $BV_{i,j}$. The value of S_i equals the sum of three adjacent pixels, while T_i equals the sum of the remaining four adjacent pixels. The input pixel value at $BV_{i,j}$ is never used in the computation.

For example, many of the linear and nonlinear spatial filtering convolution masks previously discussed were also applied to the thermal infrared image of the Savannah River and displayed in Figure 8-24. Image analysts usually work interactively with the remotely sensed data, trying different coefficients and selecting those that produce the most effective results. It is also possible to combine operators for edge detection. For example, a combination of gradient and Laplacian edge operators may yield results that are superior to

using either edge enhancement alone. In addition, nonlinear edge enhancements may provide superior results.

The Fourier Transform

Fourier analysis is a mathematical technique for separating an image into its various spatial frequency components. First, let us consider a continuous function $f(x)$. The Fourier theorem states that any function $f(x)$ can be represented by a summation of a series of sinusoidal terms of varying spatial frequencies. These terms can be obtained by the Fourier transform of $f(x)$, which is written as:

$$F(u) = \int_{-\infty}^{\infty} f(x)e^{-2\pi iux}dx , \qquad (8.23)$$

where u is spatial frequency. This means that $F(u)$ is a frequency domain function. The spatial domain function $f(x)$ can be recovered from $F(u)$ by the inverse Fourier transform:

$$f(x) = \int_{-\infty}^{\infty} F(u)e^{2\pi iux}du . \qquad (8.24)$$

To use Fourier analysis in image processing, we must consider two extensions of these equations. First, both transforms can be extended from one-dimensional functions to two-dimensional functions $f(x, y)$ and $F(u, v)$. For Equation 8.23 this becomes:

$$F(u, v) = \int\int_{-\infty}^{\infty} f(x, y)e^{-2\pi i(ux+vy)}dxdy . \qquad (8.25)$$

Furthermore, we can extend both transforms to discrete functions. The two-dimensional discrete Fourier transform is written:

$$F(u, v) = \frac{1}{NM}\sum_{x=0}^{N-1}\sum_{y=0}^{M-1} f(x, y)e^{-2\pi i\left(\frac{ux}{N}+\frac{vy}{M}\right)} \qquad (8.26)$$

where N is the number of pixels in the x direction and M is the number of pixels in the y direction. Every image can be described as a two-dimensional discrete function. Therefore, Equation 8.26 may be used to compute the Fourier transform of an image. The image can be reconstructed using the inverse transform:

$$f(x, y) = \sum_{u=0}^{N-1}\sum_{v=0}^{M-1} F(u, v)e^{2\pi i\left(\frac{ux}{N}+\frac{vy}{M}\right)} . \qquad (8.27)$$



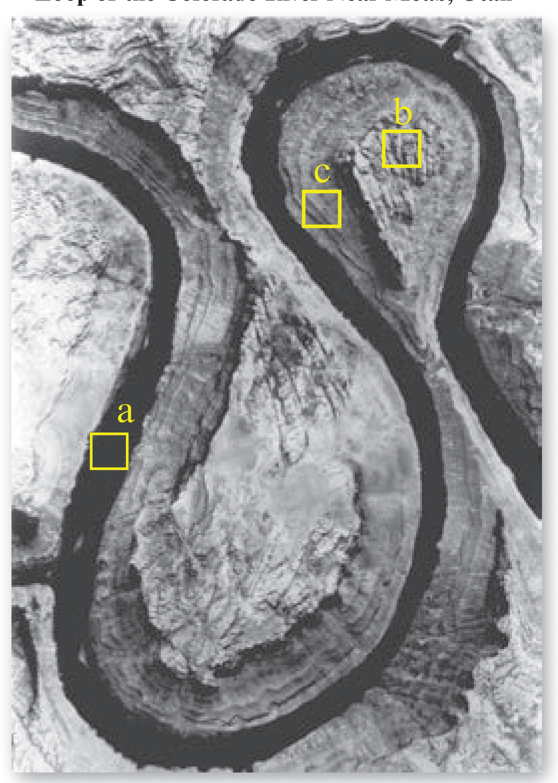


FIGURE 8–25 Digitized aerial photograph of The Loop on the Colorado River with three subimages identified.

You are probably asking the question, "What does the $F(u,v)$ represent?" It contains the spatial frequency information of the original image $f(x, y)$ and is called the *frequency spectrum.* Note that it is a complex function because it contains i, which equals $\sqrt{-1}$. We can write any complex function as the sum of a real part and an imaginary part:

$$F(u, v) = R(u, v) + iI(u, v), \qquad (8.28)$$

which is equivalent to

$$F(u, v) = |F(u, v)|e^{i\phi(u, v)} , \qquad (8.29)$$

where $|F(u, v)|$ is a real function, and

$$|F(u, v)| = \sqrt{R(u, v)^2 + I(u, v)^2} .$$

$|F(u, v)|$ is called the magnitude of the Fourier transform and can be displayed as a two-dimensional image. It represents the magnitude and the direction of the different frequency components in the image $f(x, y)$. The variable ϕ in Equation 8.29 represents *phase* information in the image $f(x, y)$. Although we usually ignore the phase information when we display the

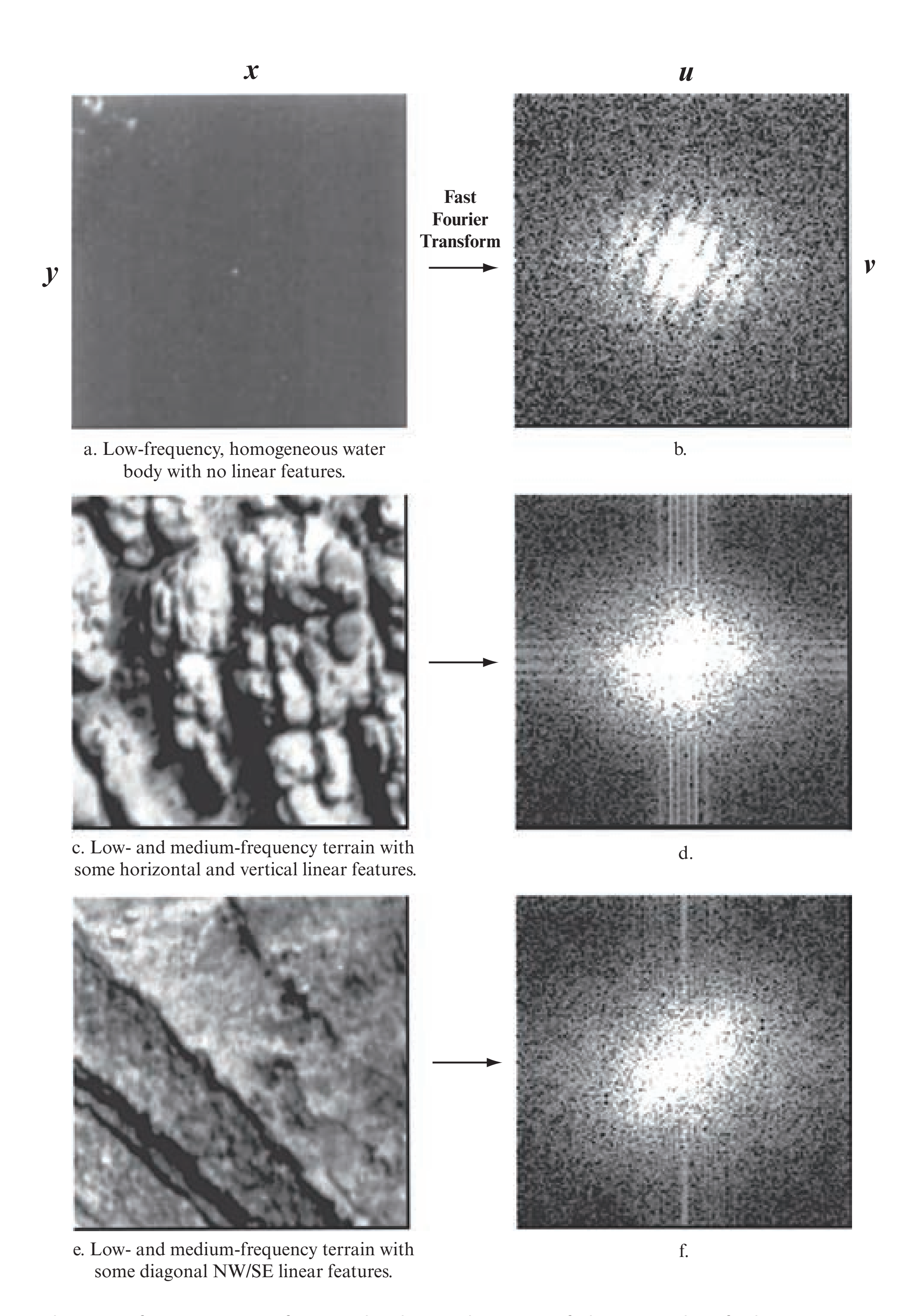


FIGURE 8–26 Application of a Fourier transform to the three subimages of The Loop identified in Figure 8-25.

Fourier transform, we cannot recover the original image without it.

To understand how the Fourier transform is useful in remote sensing applications, let us first consider three subimages extracted from The Loop, shown in Figure 8-25. The first subset includes a low-frequency water portion of the photograph shown enlarged in Figure 8-26a. Another area contains low- and medium-frequency terrain information with both horizontal and vertical linear features (Figure 8-26c). The final subset contains low- and medium-frequency terrain with some diagonal linear features (Figure 8-26e). The magnitudes of the subimages' Fourier transforms are shown in Figures 8-26b, d, and f. The Fourier magnitude images are symmetric about their centers, and *u*

Stationary Periodic Noise and Its Fourier Transform

Fast Fourier Transform →

a. High-frequency noise.

b. Fourier transform.

Fast Fourier Transform →

c. Low-frequency noise.

d. Fourier transform.

FIGURE 8–27 Two examples of stationary periodic noise and their Fourier transforms.

and v represent spatial frequency. The displayed Fourier magnitude image is usually adjusted to bring the $F(0, 0)$ to the center of the image rather than to the upper-left corner. Therefore, the intensity at the center represents the magnitude of the lowest-frequency component. The frequency increases away from the center. For example, consider the Fourier magnitude of the homogeneous water body (Figure 8-26b). The very bright values found in and around the center indicate that it is dominated by low-frequency components. In the second image, more medium-frequency components are present in addition to the background of low-frequency components. We can relatively easily identify the high-frequency information representing the horizontal and vertical linear features in the original image (Figure 8-26d). Notice the alignment of the cloud of points in the center of the Fourier transform in Figure 8-26f. It represents the diagonal linear features trending in the NW–SE direction in the photograph.

It is important to remember that the strange-looking Fourier transformed image $F(u, v)$ contains all the information found in the original image. It provides a mechanism for analyzing and manipulating images according to their spatial frequencies. It is useful for image restoration, filtering, and radiometric correction. For example, the Fourier transform can be used to remove periodic noise in remotely sensed data. When the pattern of periodic noise is unchanged throughout the image, it is called stationary periodic noise. Striping in remotely sensed imagery is usually composed of stationary periodic noise.

When stationary periodic noise is a single-frequency sinusoidal function in the spatial domain, its Fourier transform consists of a single bright point (a peak of brightness). For example, Figure 8-27a and c displays two images of sinusoidal functions with different frequencies (which look very much like striping in remote sensor data!). Figure 8-27b and d are their Fourier



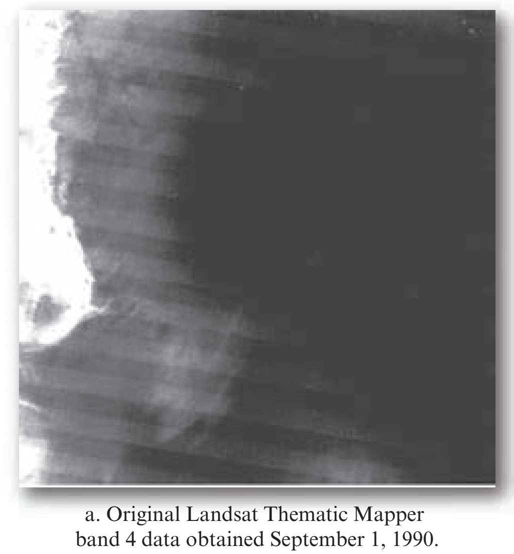
FIGURE 8–28 Landsat TM band 4 data of Al Jubail, Saudi Arabia, obtained on September 1, 1990. Image courtesy of NASA.

transforms. The frequency and orientation of the noise can be identified by the position of the bright points. The distance from the bright points to the center of the transform (the lowest-frequency component in the image) is directly proportional to the frequency. A line connecting the bright point and the center of the transformed image is always perpendicular to the orientation of the noise lines in the original image. Striping in the remotely sensed data is usually composed of sinusoidal functions with more than one frequency in the same orientation. Therefore, the Fourier transform of such noise consists of a series of bright points lined up in the same orientation.

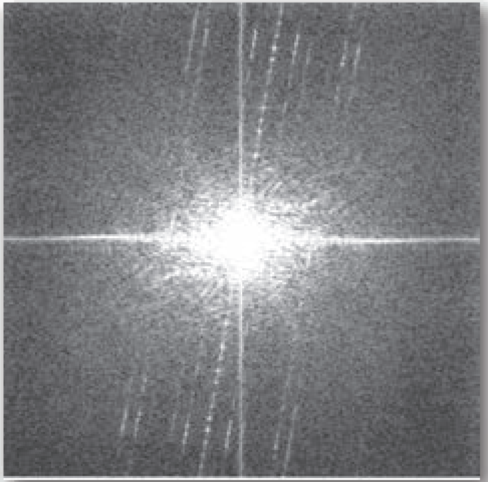
Because the noise information is concentrated in a point or a series of points in the frequency domain, it is relatively straightforward to identify and remove them in the frequency domain, whereas it is quite difficult to remove them in the standard spatial domain. Basically, an analyst can manually cut out these lines or points in the Fourier transform image or use a computer program to look for such noise and remove it. For example, consider the Landsat TM band 4 data of Al Jubail, Saudi Arabia, obtained on September 1, 1990 (Figure 8-28). The image contains serious stationary periodic striping, which can make the data unusable when conducting near-shore studies of suspended sediment transport. Figure 8-29 documents how a portion of the Landsat TM scene was corrected. First, a Fourier transform of the area was computed (Figure 8-29b). The analyst then modified the Fourier transform by selectively removing the points in the plot associated with the systematic striping (Figure 8-29c). This can be done manually or a special program can be written to look for and remove such systematic noise patterns in the Fourier transform image. The inverse Fourier transform was then computed, yielding a clean band 4 image, which may be more useful for biophysical analysis (Figure 8-29d). This type of noise could not be removed using a simple convolution mask. Rather, it requires access to the Fourier transform and selective editing out of the noise in the Fourier transform image.

Figure 8-30 depicts hyperspectral data of the Mixed Waste Management Facility (MWMF) on the Savannah River Site that was destriped using the same logic. A Fourier transform of a single band of data revealed the high-frequency information associated with the horizontal striping (Figure 8-30b). A cut filter was created to isolate the high-frequency information associated with the horizontal striping (Figure 8-30c and d). The result of applying the cut filter and inverting the fast Fourier transform is shown in Figure 8-30e. The striping was reduced substantially.

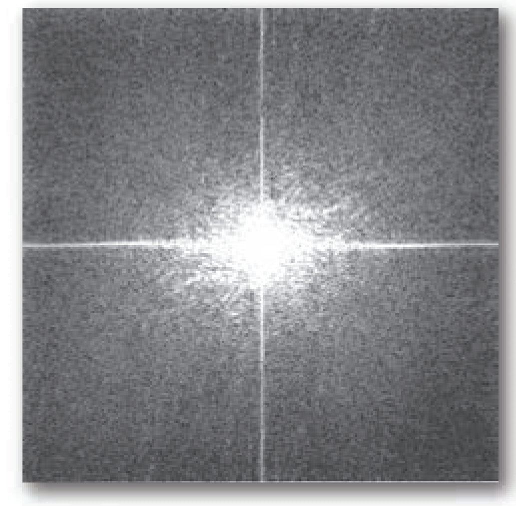
Application of a Fourier Transform to Landsat Thematic Mapper Data of an Area Near Al Jubail, Saudi Arabia, to Remove Striping



a. Original Landsat Thematic Mapper band 4 data obtained September 1, 1990.



b. Fourier transform.



c. Interactively cleaned Fourier transform.



d. Corrected Landsat Thematic Mapper band 4 data.

FIGURE 8–29 Application of a Fourier transform to a portion of Landsat TM band 4 data of Al Jubail, Saudi Arabia: a) original TM band 4 data, b) Fourier transform, c) cleaned Fourier transform, and d) destriped band 4 data.

Spatial Filtering in Frequency Domain

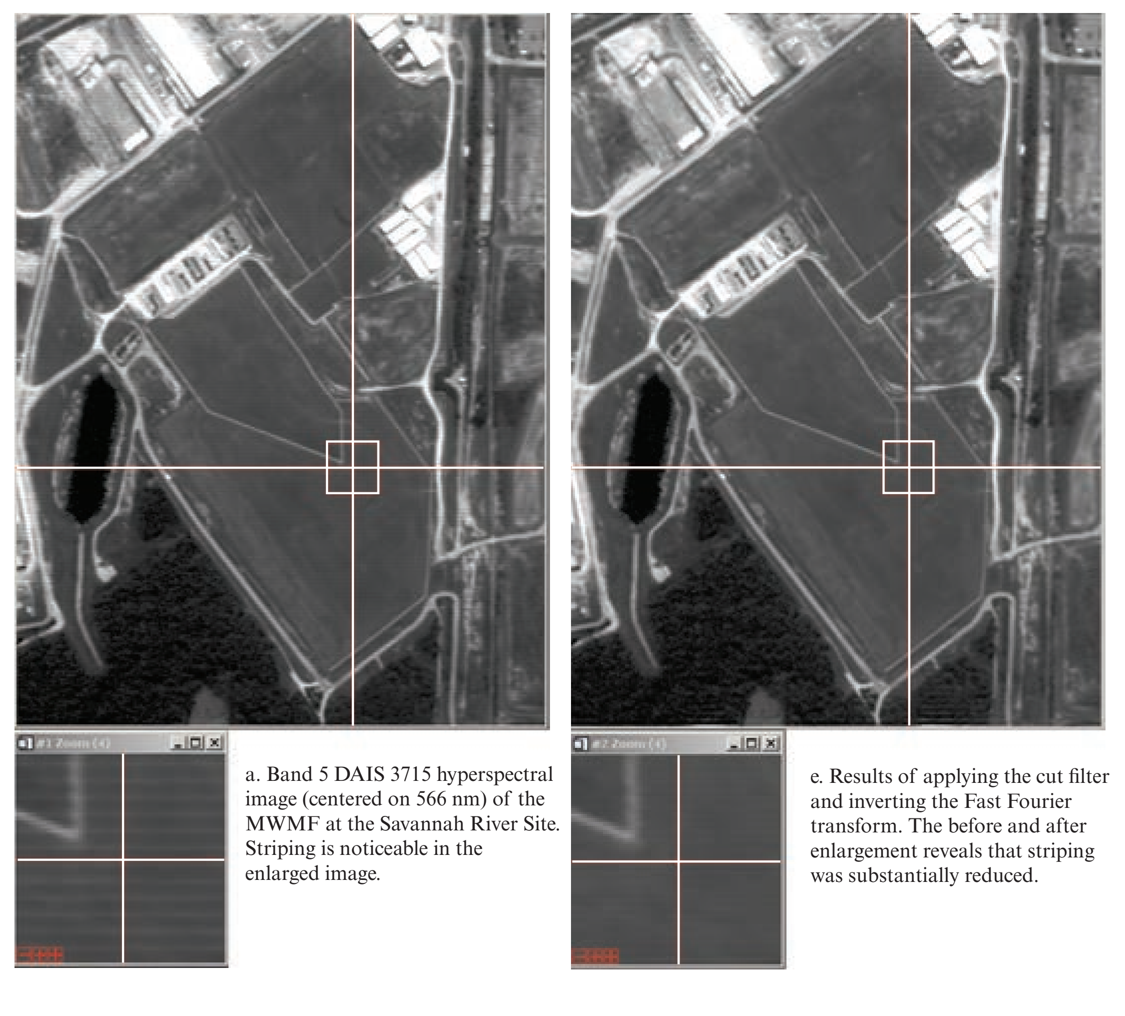
We have discussed filtering in the spatial domain using convolution filters. It can also be performed in the frequency domain. Using the Fourier transform, we can manipulate directly the frequency information of the image. The manipulation can be performed by multiplying the Fourier transform of the original image by a mask image called a frequency domain filter, which will block or weaken certain frequency components by making the values of certain parts of the frequency spectrum become smaller or even zero. Then we can compute the inverse Fourier transform of the manipulated frequency spectrum to obtain a filtered image in the spatial domain. Numerous algorithms are available for computing the Fast Fourier transform (FFT) and the inverse Fast Fourier transform (IFFT) (Russ, 2011). Spatial filtering in the frequency domain generally involves computing the FFT of the original image, multiplying the FFT of a convolution mask of the analyst's choice (e.g., a low-pass filter) with the FFT, and inverting the resultant image with the IFFT; that is,

$$f(x, y)\ \text{FFT} \rightarrow F(u, v) \rightarrow F(u, v)\ G(u, v)$$

$$\rightarrow F'(u, v)\ \text{IFFT} \rightarrow f'(x, y)\,.$$

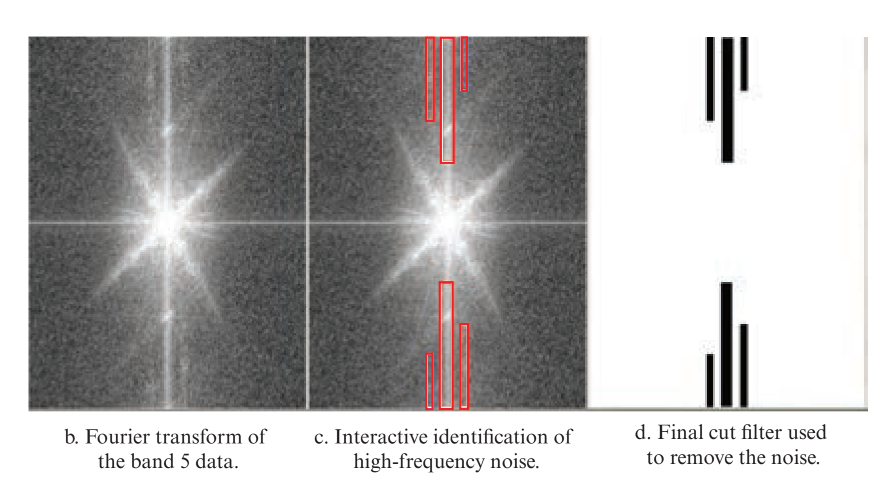
The convolution theorem states that the convolution of two images is equivalent to the multiplication of their Fourier transformations. If

$$f'(x, y) = f(x, y) * g(x, y)\,, \tag{8.30}$$



a. Band 5 DAIS 3715 hyperspectral image (centered on 566 nm) of the MWMF at the Savannah River Site. Striping is noticeable in the enlarged image.

e. Results of applying the cut filter and inverting the Fast Fourier transform. The before and after enlargement reveals that striping was substantially reduced.



b. Fourier transform of the band 5 data.

c. Interactive identification of high-frequency noise.

d. Final cut filter used to remove the noise.

FIGURE 8–30 Application of a Fourier transform to a single band of hyperspectral data of the Mixed Waste Management Facility at the Savannah River Site near Aiken, SC (based on Jensen et al., 2003).

where * represents the operation of convolution, $f(x, y)$ is the original image and $g(x,y)$ is a convolution mask filter, then

$$F'(u, v) = F(u, v)G(u, v)\,, \qquad (8.31)$$

where F', F, and G are Fourier transforms of f', f, and g, respectively.

Two examples of such manipulation are shown in Figures 8-31a and b. A low-pass filter (Mask B) and a

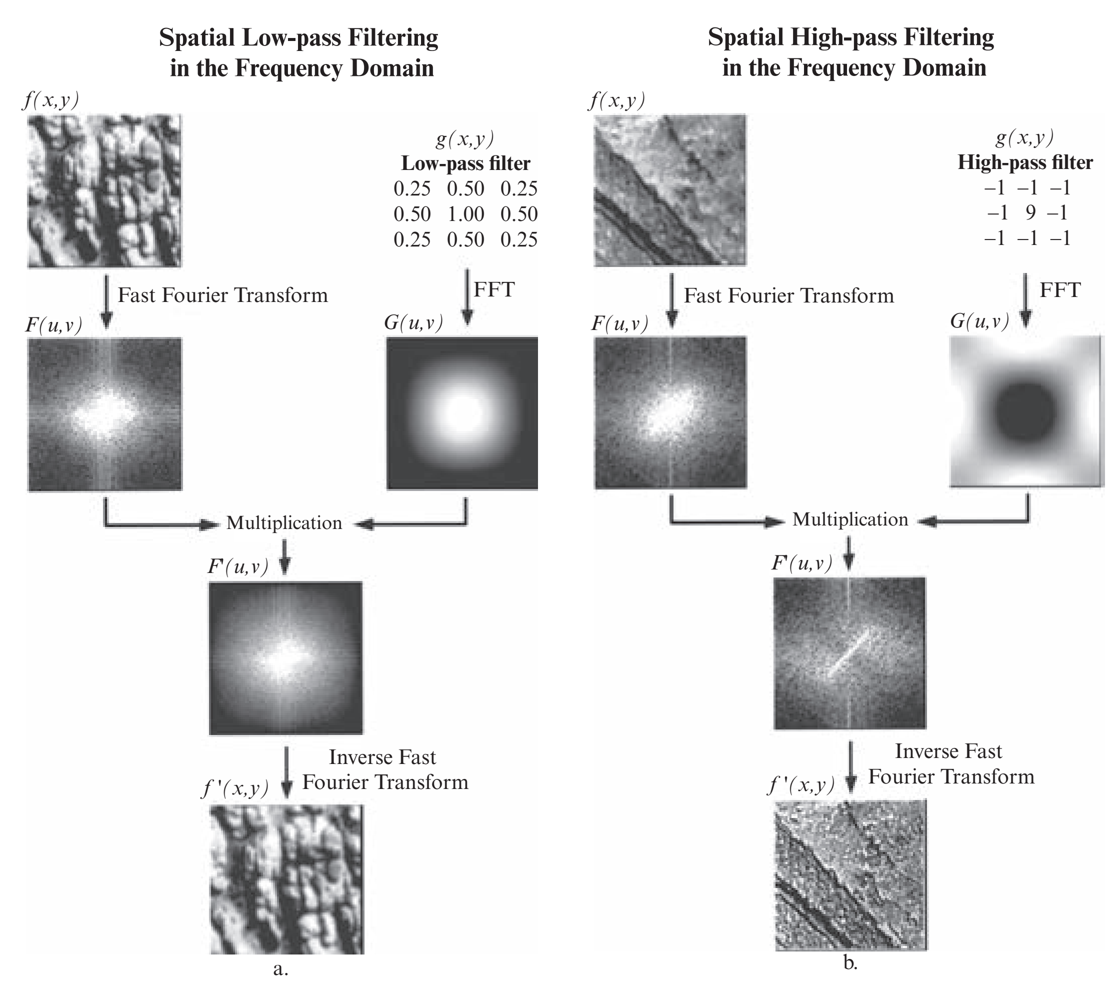
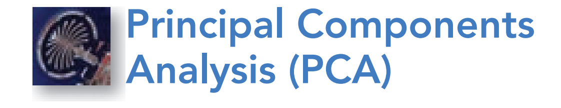


FIGURE 8–31 a) Spatial low-pass and b) high-pass filtering in the frequency domain using a Fourier transform.

high-pass filter (Mask D) were used to construct the filter function $g(x, y)$ in Figures 8-31a and b, respectively. In practice, one problem must be solved. Usually, the dimensions of $f(x, y)$ and $g(x, y)$ are different; for example, the low-pass filter in Figure 8-31a has only nine elements, while the image is composed of 128 × 128 pixels. Operation in the frequency domain requires that the sizes of $F(u, v)$ and $G(u, v)$ be the same. This means the sizes of f and g must be made the same because the Fourier transform of an image has the same size as the original image. The solution of this problem is to construct $g(x, y)$ by putting the convolution mask at the center of a zero-value image that has the same size as f. Note that in the Fourier transforms of the two convolution masks the low-pass convolution mask has a bright center (Figure 8-31a), while the high-pass filter has a dark center (Figure 8-31b). The multiplication of Fourier transforms $F(u, v)$ and $G(u, v)$ results in a new Fourier transform, $F'(u, v)$. Computing the inverse Fast Fourier transformation produces $f'(x, y)$, a filtered version of the original image. Thus, spatial filtering can be performed in both the spatial and frequency domain.'

As demonstrated, filtering in the frequency domain involves one multiplication and two transformations. For general applications, convolution in the spatial domain may be more cost-effective. Only when $g(x, y)$ is very large does the Fourier method become cost-effective. However, with the frequency domain method we can also do some filtering that is not easy to do in the spatial domain. We may construct a frequency domain filter $G(u,v)$ specifically designed to remove certain frequency components in the image.



Principal Components Analysis (PCA)

Principal components analysis (often called PCA, or Karhunen–Loeve analysis) has proven to be of great value in the analysis of multispectral and hyperspectral remotely sensed data (e.g., Mitternicht and Zinck, 2003; Amato et al., 2008; Small, 2012; Mberego et al., 2013). Principal components analysis is a technique that transforms the original remotely sensed dataset into a substantially smaller and easier to interpret set

TABLE 8–5 Charleston, SC, Thematic Mapper scene statistics used in the principal components analysis (PCA).

Band Number:	1	2	3	4	5	7	6
μm	0.45–0.52	0.52–0.60	0.63–0.69	0.76–0.90	1.55–1.75	2.08–2.35	10.4–12.5
Univariate Statistics							
Mean	64.80	25.60	23.70	27.30	32.40	15.00	110.60
Standard Deviation	10.05	5.84	8.30	15.76	23.85	12.45	4.21
Variance	100.93	34.14	68.83	248.40	568.84	154.92	17.78
Minimum	51	17	14	4	0	0	90
Maximum	242	115	131	105	193	128	130
Variance–Covariance Matrix							
1	100.93						
2	56.60	34.14					
3	79.43	46.71	68.83				
4	61.49	40.68	69.59	248.40			
5	134.27	85.22	141.04	330.71	568.84		
7	90.13	55.14	86.91	148.50	280.97	154.92	
6	23.72	14.33	22.92	43.62	78.91	42.65	17.78
Correlation Matrix							
1	1.00						
2	0.96	1.00					
3	0.95	0.96	1.00				
4	0.39	0.44	0.53	1.00			
5	0.56	0.61	0.71	0.88	1.00		
7	0.72	0.76	0.84	0.76	0.95	1.00	
6	0.56	0.58	0.66	0.66	0.78	0.81	1.00

of uncorrelated variables that represents most of the information present in the original dataset (Good et al., 2012). Principal components are extracted from the original data such that the first principal component accounts for the maximum proportion of the variance of the original dataset, and subsequent orthogonal components account for the maximum proportion of the remaining variance (Zhao and Maclean, 2000; Viscarra-Rossel and Chen, 2011). The ability to reduce the *dimensionality* (i.e., the number of bands in the dataset that must be analyzed to produce usable results) from *n* to just a few bands is an important economic consideration, especially if the potential information that can be recovered from the transformed data is just as good as the original remote sensor data.

A form of PCA may also be useful for reducing the dimensionality of hyperspectral datasets. For example, many scientists use a *minimum noise fraction* (MNF) procedure, which is a cascaded principal components analysis for data compression and noise reduction of hyperspectral data (Jensen et al., 2003). Chapter 11 provides an example of the use of the MNF procedure.

To perform principal components analysis we apply a transformation to a *correlated* set of multispectral data. For example, the Charleston TM scene is a likely candidate because bands 1, 2, and 3 are highly correlated, as are bands 5 and 7 (Table 8-5). The application of the transformation to the correlated remote sensor data will result in an *uncorrelated* multispectral dataset

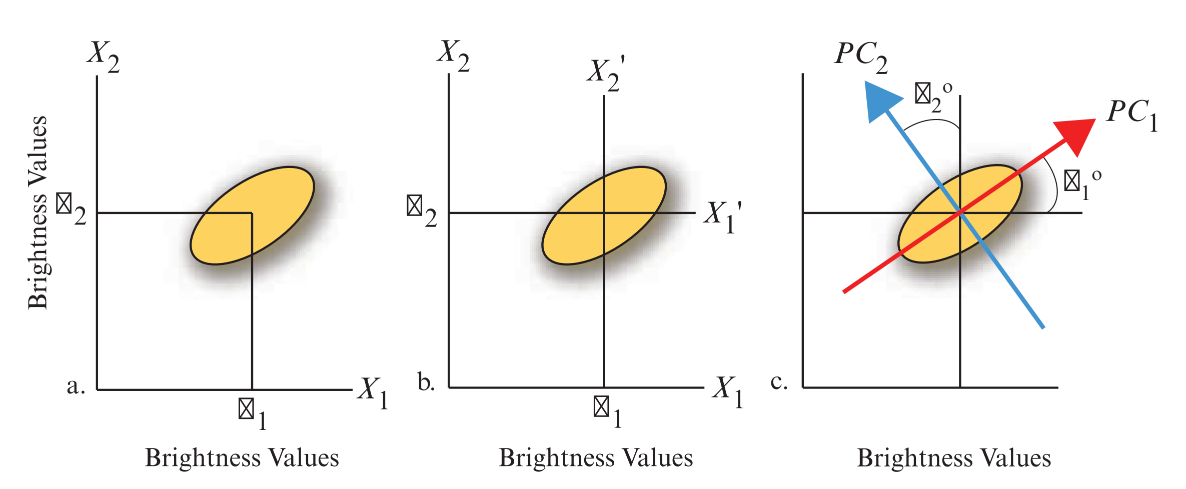


FIGURE 8–32 Diagrammatic representation of the spatial relationship between the first two principal components. a) Scatterplot of data points collected from two remotely sensed bands labeled X_1 and X_2 with the means of the distribution labeled μ_1 and μ_2. b) A new coordinate system is created by shifting the axes to an X' system. The values for the new data points are found by the relationship $X_1{}' = X_1 - \mu_1$ and $X_2{}' = X_2 - \mu_2$. c) The X' axis system is then rotated about its origin (μ_1, μ_2) so that PC_1 is projected through the semi-major axis of the distribution of points and the variance of PC_1 is a maximum. PC_2 must be perpendicular to PC_1. The PC axes are the principal components of this two-dimensional data space. Principal component 1 often accounts for > 90% of the variance, and principal component 2 accounts for 2% to 10%, etc.

that has certain ordered variance properties. This transformation is conceptualized by considering the two-dimensional distribution of pixel values obtained in two TM bands, which we will label X_1 and X_2. A scatterplot of all the brightness values associated with each pixel in each band is shown in Figure 8-32a, along with the location of the respective means, μ_1 and μ_2. The spread or variance of the distribution of points is an indication of the correlation and quality of information associated with both bands. If all the data points were clustered in an extremely tight region in the two-dimensional space, these data would probably provide very little information.

The initial measurement coordinate axes (X_1 and X_2) may not be the best arrangement in multispectral feature space to analyze the remote sensor data associated with these two bands. The goal is to use principal components analysis to *translate* and/or *rotate* the original axes so that the original brightness values on axes X_1 and X_2 are redistributed (reprojected) onto a new set of axes or dimensions, X'_1 and X'_2. For example, the best *translation* for the original data points from X_1 to X'_1 and from X_2 to X'_2 coordinate systems might be the simple relationship $X'_1 = X_1 - \mu_1$ and $X'_2 = X_2 - \mu_2$. Thus, the new coordinate system origin (X'_1 and X'_2) now lies at the location of both means in the original scatter of points (Figure 8-32b). The X' coordinate system might then be *rotated* about its new origin (μ_1, μ_2) in the new coordinate system some ϕ degrees so that the first axis X'_1 is associated with the maximum amount of variance in the scatter of points (Estes et al., 2010) (Figure 8-32c). This new axis is called the first *principal component* ($PC_1 = \lambda_1$). The second principal component ($PC_2 = \lambda_2$) is perpendicular (orthogonal) to PC_1. Thus, the major and minor axes of the ellipsoid of points in bands X_1 and X_2 are called the principal components. The third, fourth, fifth, and so on, components contain decreasing amounts of the variance found in the dataset.

To transform (reproject) the original data on the X_1 and X_2 axes onto the PC_1 and PC_2 axes, we must obtain certain transformation coefficients that we can apply in a linear fashion to the original pixel values. The linear transformation required is derived from the covariance matrix of the original dataset. Thus, this is a data-dependent process with each new remote sensing dataset yielding different transformation coefficients.

The transformation is computed from the original spectral statistics, as follows (Short, 1982):

1. The $n \times n$ *covariance matrix*, *Cov*, of the n-dimensional remote sensing dataset to be transformed is computed (Table 8-5). Use of the covariance matrix results in a non-standardized PCA, whereas use of the correlation matrix results in a standardized PCA (Eastman and Fulk, 1993; Carr, 1998).

The eigenvalues, $E = [\lambda_{1,1}, \lambda_{2,2}, \lambda_{3,3}, \ldots, \lambda_{n,n}]$, and eigenvectors, $EV = [a_{kp} \ldots$ for $k = 1$ to n bands, and $p = 1$ to n components], of the covariance matrix are computed such that:

TABLE 8–6 Eigenvalues computed for the covariance matrix.

	Component *p*						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Eigenvalues, λ_p	1,010.92	131.20	37.60	6.73	3.95	2.17	1.24
Difference	879.72	93.59	30.88	2.77	1.77	.93	--
Total Variance = 1193.81							

Percent of total variance in the data explained by each component:

Computed as $\%_p = \dfrac{\text{eigenvalue } \lambda_p \times 100}{\sum_{p=1}^{7} \text{eigenvalue } \lambda_p}$.

For example,

$$\sum_{p=1}^{7} \lambda_p = 1{,}010.92 + 131.20 + 37.60 + 6.73 + 3.95 + 2.17 + 1.24 = 1{,}193.81 .$$

Percentage of variance explained by first component $= \dfrac{1{,}010.92 \times 100}{1{,}193.81} = 84.68$.

	1	**2**	**3**	**4**	**5**	**6**	**7**
Percentage	84.68	10.99	3.15	0.56	0.33	0.18	0.10
Cumulative	84.68	95.67	98.82	99.38	99.71	99.89	100.00

$$EV \cdot Cov \cdot EV^T = \overset{E}{\begin{bmatrix} \lambda_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_{3,3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{4,4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_{5,5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_{6,6} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{n,n} \end{bmatrix}} \quad (8.32)$$

where EV^T is the transpose of the eigenvector matrix, EV, and E is a diagonal covariance matrix whose elements λ_{ii}, called *eigenvalues*, are the variances of the *p*th *principal components*, where $p = 1$ to n components. The nondiagonal eigenvalues, λ_{ij}, are equal to zero and therefore can be ignored. The number of nonzero eigenvalues in an $n \times n$ covariance matrix always equals n, the number of bands examined. The eigenvalues are often called *components* (i.e., eigenvalue 1 may be referred to as principal component 1). Eigenvalues and eigenvectors were computed for the Charleston, SC, Thematic Mapper scene (Tables 8-6 and 8-7).

The eigenvalues contain important information. For example, it is possible to determine the percent of total variance explained by each of the principal components, $\%_p$, using the equation:

$$\%_p = \frac{\text{eigenvalue } \lambda_p \times 100}{\sum_{p=1}^{n} \text{eigenvalue } \lambda_p}, \quad (8.33)$$

where λ_p is the *p*th eigenvalue out of the possible n eigenvalues. For example, the first principal component (eigenvalue λ_1) of the Charleston TM scene accounts for 84.68% of the variance in the entire multispectral dataset (Table 8-6). Component 2 accounts for 10.99% of the remaining variance. Cumulatively, these first two principal components account for 95.67% of the variance. The third component accounts for another 3.15%, bringing the total to 98.82% of the variance explained by the first three components (Table 8-6). Thus, the seven-band TM dataset of Charleston might be compressed into just three new principal component images (or bands) that explain 98.82% of the variance.

But what do these new components represent? For example, what does component 1 stand for? By computing the correlation of each band k with each component p, it is possible to determine how each

TABLE 8–7 Eigenvectors (a_{kp}) (factor scores) computed for the covariance matrix found in Table 8-5.

	Component p						
	1	**2**	**3**	**4**	**5**	**6**	**7**
band$_k$ 1	0.205	0.637	0.327	–0.054	0.249	–0.611	–0.079
2	0.127	0.342	0.169	–0.077	0.012	0.396	0.821
3	0.204	0.428	0.159	–0.076	–0.075	0.649	–0.562
4	0.443	–0.471	0.739	0.107	–0.153	–0.019	–0.004
5	0.742	–0.177	–0.437	–0.300	0.370	0.007	0.011
7	0.376	0.197	–0.309	–0.312	–0.769	–0.181	0.051
6	0.106	0.033	–0.080	0.887	0.424	0.122	0.005

band “loads” or is associated with each principal component. The equation is:

$$R_{kp} = \frac{a_{kp} \times \sqrt{\lambda_p}}{\sqrt{\mathrm{Var}_k}} \tag{8.34}$$

where

a_{kp} = eigenvector for band k and component p
λ_p = pth eigenvalue
Var_k = variance of band k in the covariance matrix.

This computation results in a new n ⊠ n matrix (Table 8-8) filled with *factor loadings.* For example, the highest correlations (i.e., factor loadings) for principal component 1 were for bands 4, 5, and 7 (0.894, 0.989, and 0.961, respectively; Table 8-8). This suggests that this component is a near- and middle-infrared reflectance band. This makes sense because the golf courses and other vegetation are particularly bright in this image (Figure 8-33). Conversely, principal component 2 has high loadings only in the visible bands 1, 2, and 3 (0.726, 0.670, and 0.592), and vegetation is noticeably darker in this image (Figure 8-33). This is a visible spectrum component. Component 3 loads heavily on the near-infrared (0.287) and appears to provide some unique vegetation information. Component 4 accounts for little of the variance but is easy to label since it loads heavily (0.545) on the thermal-infrared band 6. Components 5, 6, and 7 provide no useful information and contain most of the systematic noise. They account for very little of the variance and should probably not be used further.

Now that we understand what information each component contributes, it is useful to see how the principal component images are created. To do this, it is necessary to first identify the original brightness values ($BV_{i,j,k}$) associated with a given pixel. In this case we will evaluate the first pixel in a hypothetical image at row 1, column 1 for each of seven bands. We will represent this as the vector X, such that

$$X = \begin{bmatrix} BV_{1,1,1} = 20 \\ BV_{1,1,2} = 30 \\ BV_{1,1,3} = 22 \\ BV_{1,1,4} = 60 \\ BV_{1,1,5} = 70 \\ BV_{1,1,7} = 62 \\ BV_{1,1,6} = 50 \end{bmatrix}$$

We will now apply the appropriate transformation to this data such that they are projected onto the first principal component’s axes. In this way we will find out what the new brightness value (new $BV_{i,j,p}$) will be for this component, p. It is computed using the equation:

$$\mathrm{new}BV_{i,j,p} = \sum_{k=1}^{n} a_{kp} BV_{i,j,k} \tag{8.35}$$

where a_{kp} = eigenvectors, $BV_{i,j,k}$ = brightness value in band k for the pixel at row i, column j, and n = number of bands. In our hypothetical example, this yields:

$$\begin{aligned} \mathrm{new}BV_{1,1,1} &= a_{1,1}(BV_{1,1,1}) + a_{2,1}(BV_{1,1,2}) + \\ &\quad a_{3,1}(BV_{1,1,3}) + a_{4,1}(BV_{1,1,4}) + a_{5,1}(BV_{1,1,5}) + \\ &\quad a_{6,1}(BV_{1,1,7}) + a_{7,1}(BV_{1,1,6}) \\ &= 0.205(20) + 0.127(30) + 0.204(22) + 0.443(60) + \\ &\quad 0.742(70) + 0.376(62) + 0.106(50) \\ &= 119.53 \end{aligned}$$

TABLE 8–8 Degree of correlation, R_{kp}, between each band k and each principal component p.

Computed: $$R_{kp} = \frac{a_{kp} \times \sqrt{\lambda_p}}{\sqrt{\mathrm{Var}_k}}.$$

For example:

$$R_{1,1} = \frac{0.205 \times \sqrt{1{,}010.92}}{\sqrt{100.93}} = \frac{0.205 \times 31.795}{10.046} = 0.649$$

$$R_{5,1} = \frac{0.742 \times \sqrt{1{,}010.92}}{\sqrt{568.84}} = \frac{0.742 \times 31.795}{23.85} = 0.989$$

$$R_{2,2} = \frac{0.342 \times \sqrt{131.20}}{\sqrt{34.14}} = \frac{0.342 \times 11.45}{5.842} = 0.670.$$

	Component p						
	1	2	3	4	5	6	7
Band$_k$ 1	0.649	0.726	0.199	–0.014	0.049	–0.089	–0.008
2	0.694	0.670	0.178	–0.034	0.004	0.099	0.157
3	0.785	0.592	0.118	–0.023	–0.018	0.115	–0.075
4	0.894	–0.342	0.287	0.017	–0.019	–0.002	–0.000
5	0.989	–0.084	–0.112	–0.032	0.030	0.000	0.000
7	0.961	0.181	–0.152	0.065	–0.122	–0.021	0.004
6	0.799	0.089	–0.116	0.545	0.200	0.042	0.001

This pseudo-measurement is a linear combination of original brightness values and factor scores (eigenvectors). The new brightness value for row 1, column 1 in principal component 1 after truncation to an integer is new $BV_{1,1,1} = 119$.

This procedure takes place for every pixel in the original remote sensor image data to produce the principal component 1 image dataset. Then p is incremented by 1 and principal component 2 is created pixel by pixel. This is the method used to produce the principal component images shown in Figure 8-33a–g. If desired, any two or three of the principal components can be placed in the blue, green, and/or red graphics display channels to create a principal component color composite. For example, principal components 1, 2, and 3 are displayed as an RGB color composite in Figure 8-33i. These displays often depict more subtle differences in color shading and distribution than in the traditional color-infrared color composite image.

If components 1, 2, and 3 account for a significant amount of the variance in the dataset, perhaps the original seven bands of TM data can be set aside, and the remainder of the image enhancement or classification can be performed using just these three principal component images. This reduces the amount of data to be analyzed and completely bypasses the expensive and time-consuming process of feature (band) selection so often necessary when classifying remotely sensed data (discussed in Chapter 9).

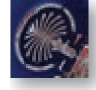
Eastman and Fulk (1993) suggested that *standardized PCA* (based on the computation of eigenvalues from correlation matrices) is superior to non-standardized PCA (computed from covariance matrices) when analyzing change in multitemporal image datasets. They processed 36 monthly AVHRR-derived normalized difference vegetation index (NDVI) images of Africa for the years 1986 to 1988. They found the first component was always highly correlated with NDVI regardless of season, while the second, third, and fourth components related to seasonal changes in NDVI. Mitternicht and Zinck (2003) found PCA very useful for delineating saline from nonsaline soils. Amato et al. (2008) used PCA and independent component analysis (ICA) to perform cloud detection using European Meteosat data. Good et al. (2012) used PCA in their development of the ATSR Saharan Dust Index (ASDI) based on the use of thermal infrared imagery from the Along-Track Scanning Radiometers (ATSRs). Interestingly, the first component provided clear-sky information while the second component was found to be the ASDI. Estes et al. (2010) developed a PCA-derived

Principal Component Images of Charleston, SC, Derived from Landsat Thematic Mapper Imagery Obtained on November 9, 1982

a. Principal component 1.
b. Principal component 2.
c. Principal component 3.
d. Principal component 4.
e. Principal component 5.
f. Principal component 6.
g. Principal component 7.
h. RGB = Landsat TM bands 4,3,2.
i. RGB = PCA bands 1,2,3.

FIGURE 8–33 a-g) Seven principal component images of the Charleston Thematic Mapper data computed using all seven bands. Component 1 consists of both near- and middle-infrared information (bands 4, 5, and 7). Component 2 contains primarily visible light information (bands 1, 2, and 3). Component 3 contains primarily near-infrared information. Component 4 consists of the thermal infrared information contributed by band 6. Thus, the seven-band TM data can be reduced in dimension to just four principal components (1, 2, 3, and 4), which account for 99.38% of the variance. h) A color-infrared color composite consisting of the original Landsat TM bands 4, 3, and 2 is displayed. i) A color composite of principal component bands 1, 2, and 3. Original imagery courtesy of NASA.

Canopy Structure Index (CSI) for modeling habitat structural complexity. Viscarra-Rosel and Chen (2011) used PCA to analyze visible and near-infrared spectra and other environmental variables to map soils in Australia. Small (2012) used PCA and temporal mixture modeling to perform time-space characterization of multiple dates of MODIS Enhanced Vegetation Index (EVI) data (discussed in the next section) of the Ganges Brahmaputra delta in India.



Vegetation Indices (VI)

Approximately 70% of the Earth's land surface is covered with vegetation. Knowledge about variations in vegetation species and community distribution patterns, alterations in vegetation phenological (growth) cycles, and modifications in the plant physiology and morphology provides valuable insight into the climatic,

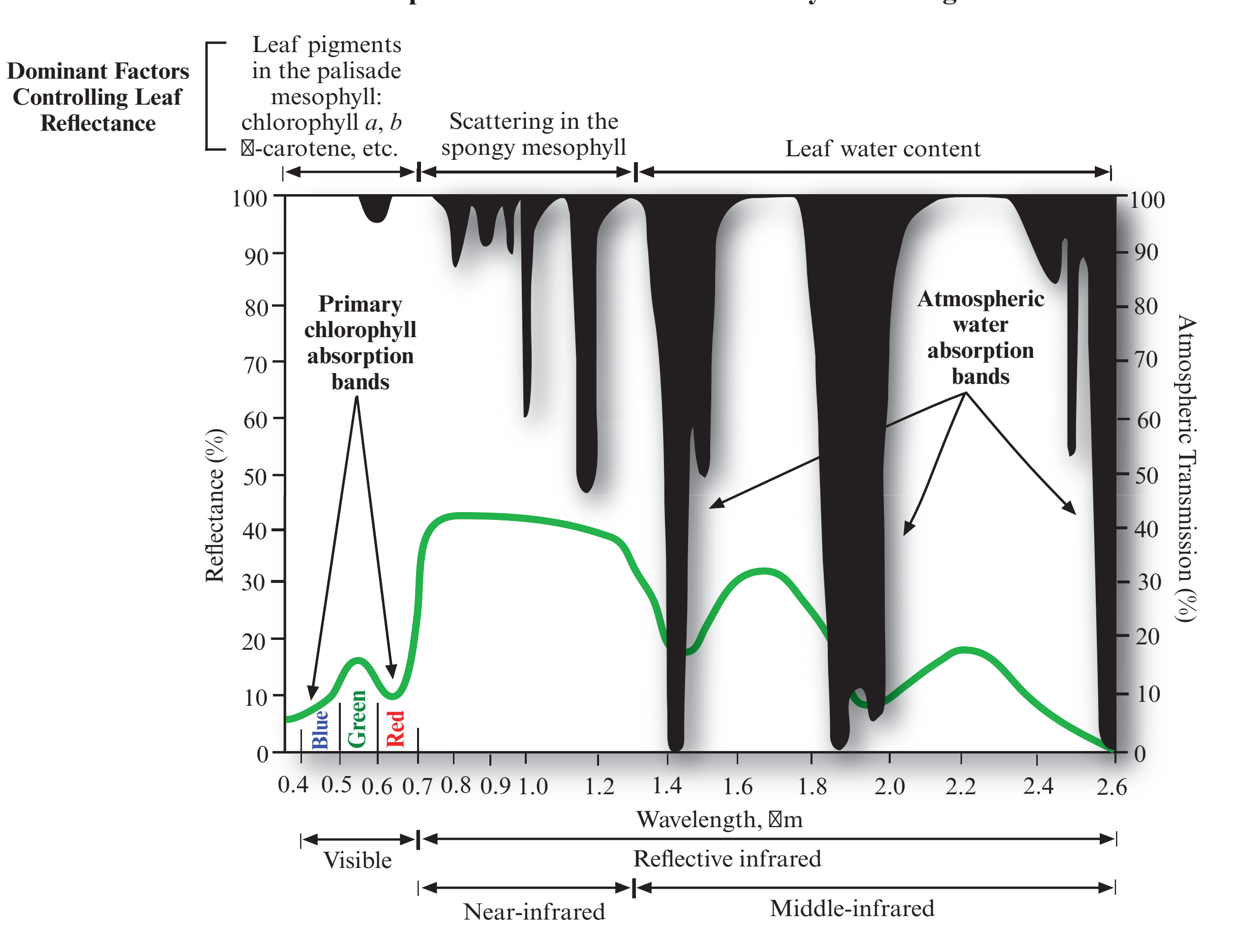


FIGURE 8–34 Spectral reflectance characteristics of healthy, green vegetation for the wavelength interval 0.4 – 2.6 μm. The dominant factors controlling leaf reflectance are the various leaf pigments in the palisade mesophyll (e.g., chlorophyll *a* and *b*, and β-carotene), the scattering of near-infrared energy in the spongy mesophyll, and the amount of water in the plant. The primary chlorophyll absorption bands occur at 0.43 – 0.45 μm and 0.65 – 0.66 μm in the visible region. The primary water atmospheric absorption bands occur at 0.97, 1.19, 1.45, 1.94, and 2.7 μm.

edaphic, geologic, and physiographic characteristics of an area (e.g., Jackson and Jensen, 2005; Im and Jensen, 2009). Furthermore, the goal of global agriculture production and the grain sector of most economies is to feed billions of people. Therefore, it is not surprising that scientists have devoted significant effort to develop remote sensing systems and visual and digital image processing algorithms to extract important vegetation biophysical information from remotely sensed data. Many of the remote sensing techniques are generic in nature and may be applied to a variety of vegetated landscapes, including:

- agriculture,
- forests,
- rangeland,
- wetland, and
- manicured urban vegetation.

This section introduces the fundamental concepts associated with vegetation biophysical characteristics and how remotely sensed data can be processed to provide unique information about these parameters. It then summarizes some of the vegetation indices developed to extract biophysical vegetation information from digital remote sensor data.

Before presenting the various vegetation indices, however, it is important to understand how the indices are related to various leaf physiological properties. Therefore, we will first discuss the dominant factors controlling leaf reflectance (Figure 8-34). This will help us appreciate why the linear combination of certain bands can be used in vegetation indices as a surrogate for leaf and/or canopy biophysical properties.

Dominant Factors Controlling Leaf Reflectance

Pioneering work by Gates et al. (1965), Gausmann et al. (1969), Myers (1970) and others demonstrated the importance of understanding how leaf pigments, internal scattering, and leaf water content affect the reflectance and transmittance properties of leaves (Peterson and Running, 1989). Dominant factors controlling leaf reflectance in the region from 0.35 – 2.6 μm are summarized in Figure 8-34.

Visible Light Interaction with Pigments in the Palisade Mesophyll Cells

Photosynthesis is an energy-storing process that takes place in leaves and other green parts of plants in the presence of light. The photosynthetic process is:

$$6CO_2 + 6H_2O + \text{light energy} \rightarrow C_6H_{12}O_6 + 6O_2 .$$

Sunlight provides the energy that powers photosynthesis. The light energy is stored in a simple sugar molecule (glucose) that is produced from carbon dioxide (CO_2) present in the air and water (H_2O) absorbed by the plant primarily through the root system. When the carbon dioxide and the water are combined and form a sugar molecule ($C_6H_{12}O_6$) in a chloroplast, oxygen gas (O_2) is released as a by-product. The oxygen diffuses out into the atmosphere.

The photosynthetic process begins when sunlight strikes *chloroplasts*, which are small bodies in the leaf that contain a green substance called chlorophyll. It is the process of food-making via photosynthesis that determines how a leaf and the associated plant canopy components appear on remotely sensed images.

Plants have adapted their internal and external structure to perform photosynthesis. This structure and its interaction with electromagnetic energy have a direct impact on how leaves and canopies appear spectrally when recorded using remote sensing instruments.

The leaf is the primary photosynthesizing organ. A cross-section of a typical green leaf is shown in Figure 8-35. The cell structure of leaves is highly variable depending upon species and environmental conditions during growth. Carbon dioxide enters the leaf from the atmosphere through tiny pores called stomata or stoma, located primarily on the underside of the leaf on the lower epidermis. Each stomata is surrounded by guard cells that swell or contract. When they swell, the stomata pore opens and allows carbon dioxide to enter the leaf. A typical sunflower leaf might have 2 million stomata, but they make up only about 1 percent of the leaf's surface area. Usually, there are more stomata on the bottom of a leaf; however, on some leaves the stomata are evenly distributed on both the upper and lower epidermis.

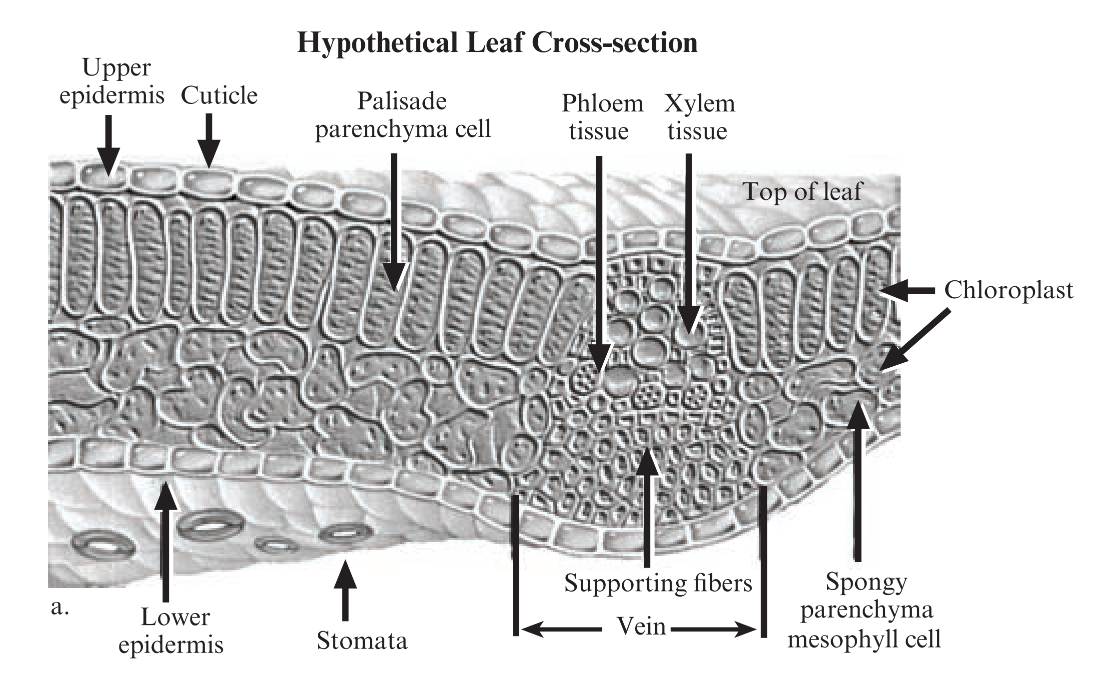
The top layer of leaf *upper epidermis* cells has a *cuticular* surface that diffuses but reflects very little light. It is variable in thickness but is often only 3 – 5 μm thick with cell dimensions of approximately 18 ×15 × 20 μm. It is useful to think of it as a waxy, translucent material similar to the cuticle at the top of your fingernail. Leaves of many plants that grow in bright sunlight have a thick cuticle that can filter out some light and guard against excessive plant water loss. Conversely, some plants such as ferns and some shrubs on the forest floor must survive in shaded conditions. The leaves of many of these plants have a thin cuticle so that the plant can collect as much of the dim sunlight as possible for photosynthesis.

Many leaves in direct sunlight have hairs growing out of the upper (and lower) epidermis, causing them to feel fuzzy. These hairs can be beneficial, as they reduce the intensity of the incident sunlight to the plant. Nevertheless, much of the visible and near-infrared wavelength energy is transmitted through the cuticle and upper epidermis to the palisade parenchyma mesophyll cells and spongy parenchyma mesophyll cells below.

Photosynthesis occurs inside the typical green leaf in two kinds of food-making cells—*palisade parenchym*a and *spongy parenchyma* mesophyll cells. Most leaves have a distinct layer of long palisade parenchyma cells in the upper part of the mesophyll and more irregularly shaped, loosely arranged spongy parenchyma cells in the lower part of the mesophyll. The palisade cells tend to form in the portion of the mesophyll toward the side from which the light enters the leaf. In most horizontal (planophile) leaves the palisade cells will be toward the upper surface, but in leaves that grow nearly vertical (erectophile), the palisade cells may form from both sides. In some leaves the elongated palisade cells are entirely absent and only spongy parenchyma cells will exist within the mesophyll.

The cellular structure of the leaf is large compared to the wavelengths of light that interact with it. Palisade cells are typically 15 × 15 × 60 μm, while spongy mesophyll parenchyma cells are smaller. The palisade parenchyma mesophyll plant cells contain chloroplasts with chlorophyll pigments.

The chloroplasts are generally 5 – 8 μm in diameter and about 1 μm in width. As many as 50 chloroplasts may be present in each parenchyma cell. Within the chloroplasts there are long, slender *grana* strands (not shown in Figure 8-35) within which the chlorophyll is actually located (approximately 0.5 μm in length and 0.05 μm in diameter). The chloroplasts are generally



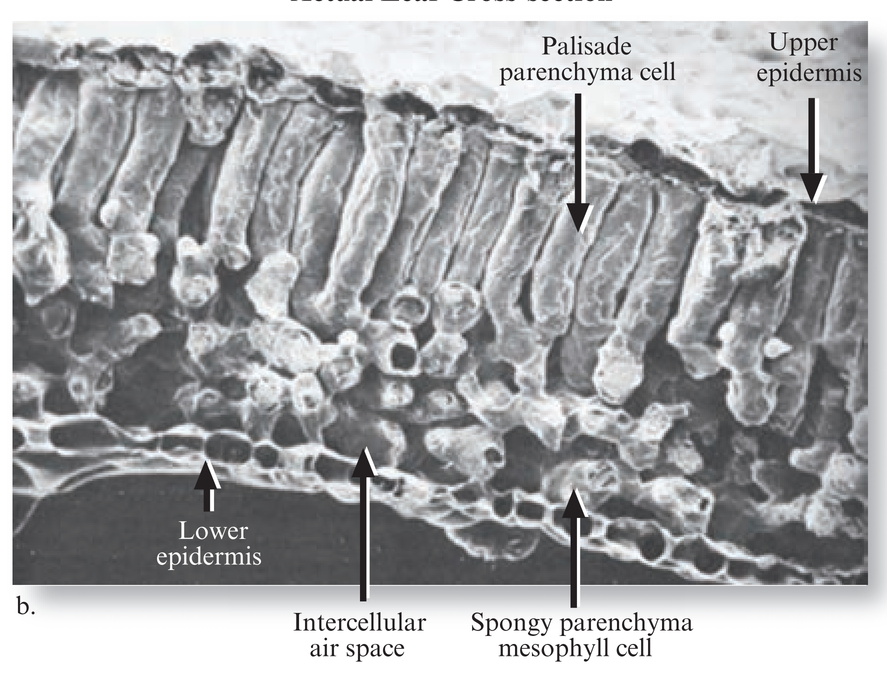


FIGURE 8–35 a) Hypothetical cross-section of a healthy green leaf showing both the top and underside of the leaf. The chlorophyll pigments in the palisade parenchyma cells have a significant impact on the absorption and reflectance of visible light (blue, green, and red), while the spongy parenchyma mesophyll cells have a significant impact on the absorption and reflectance of near-infrared incident energy. b) Electron microscope image of a green leaf.

more abundant toward the upper side of the leaf in the palisade cells and hence account for the darker green appearance of the upper leaf surface compared with the bottom lighter surface.

A molecule, when struck by a photon of light, reflects some of the energy or it can absorb the energy and thus enter into a higher energy or excited state (refer to Chapter 6). Each molecule absorbs or reflects its own characteristic wavelengths of light. Molecules in a typical green plant have evolved to absorb wavelengths of light in the visible region of the spectrum (0.35 – 0.70 μm) very well and are called *pigments*. An *absorption spectrum* for a particular pigment describes the wavelengths at which it can absorb light and enter into an excited state. Figure 8-36a presents the absorption

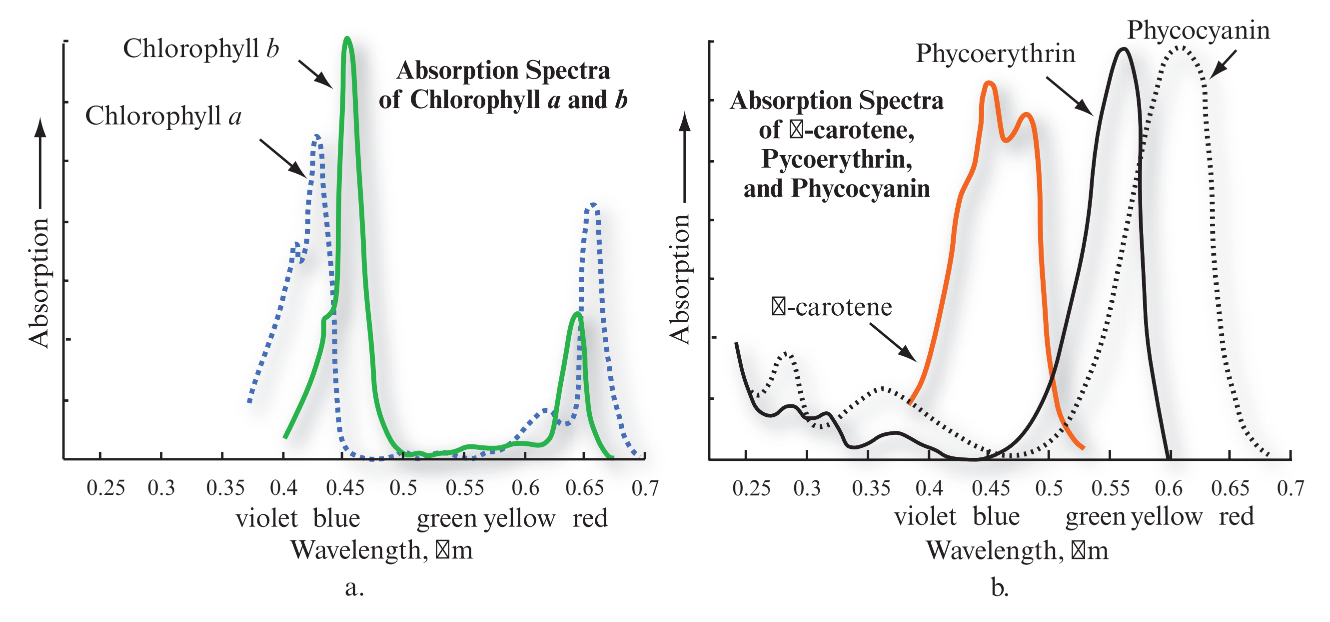


FIGURE 8–36 a) Absorption spectra of chlorophyll *a* and *b* pigments. Chlorophyll *a* and *b* pigments in the leaf absorb much of the incident blue and red wavelength energy: chlorophyll *a* at wavelengths centered on 0.43 and 0.66 μm and chlorophyll *b* at wavelengths centered on 0.45 and 0.65 μm. b) Absorption spectra of β -carotene, which absorbs primarily in the blue has a strong absorption band centered at about 0.45 μm. Other pigments that might be found in a leaf include Phycoerythrin which absorbs primarily green light, and Phycocyanin which absorbs primarily green and red light.

spectrum of pure chlorophyll pigments in solution. Chlorophyll *a* and *b* are the most important plant pigments absorbing blue and red light: chlorophyll *a* at wavelengths of 0.43 and 0.66 μm and chlorophyll *b* at wavelengths of 0.45 and 0.65 μm. A relative lack of absorption in the wavelengths between the two chlorophyll absorption bands produces a trough in the absorption efficiency at approximately 0.54 μm in the green portion of the electromagnetic spectrum (Figure 8-36a). Thus, it is the relatively lower absorption of green wavelength light (compared to blue and red light) by the leaf that causes healthy green foliage to appear green to our eyes.

There are other pigments present in the palisade mesophyll cells that are usually masked by the abundance of chlorophyll pigments. For example, there are yellow carotenes and pale yellow *xanthophyll* pigments, with strong absorption primarily in the blue wavelength region. The β -carotene absorption spectra is shown in Figure 8-36b with its strong absorption band centered at about 0.45 μm. *Phycoerythrin* pigments may also be present in the leaf, which absorb predominantly in the green region centered at about 0.55 μm, allowing blue and red light to be reflected. *Phycocyanin* pigments absorb primarily in the green and red regions centered at about 0.62 μm, allowing much of the blue and some of the green light (i.e., the combination produces cyan) to be reflected (Figure 8-36b). Because chlorophyll *a* and *b* chloroplasts are also present and have a similar absorption band in this blue region, they tend to dominate and mask the effect of the other pigments present. When a plant undergoes senescence in the fall or encounters stress, the chlorophyll pigment may disappear, allowing the carotenes and other pigments to become dominant. For example, in the fall, chlorophyll production ceases, causing the yellow coloration of the carotenes and other specific pigments in the tree foliage to become more visible to our eyes. In addition, some trees produce great quantities of *anthocyanin* in the fall, causing the leaves to appear bright red.

The two optimum spectral regions for sensing the chlorophyll absorption characteristics of a leaf are believed to be 0.45 – 0.52 μm and 0.63 – 0.69 μm (Figure 8-36a). The former region is characterized by strong absorption by carotenoids and chlorophylls, whereas the latter is characterized by strong chlorophyll absorption. Remote sensing of chlorophyll absorption within a canopy represents a fundamental biophysical variable useful for many biogeographical investigations. The absorption characteristics of plant canopies may be coupled with other remotely sensed data to identify vegetation stress, yield, and other hybrid variables. Thus, many remote sensing studies are concerned with monitoring what happens to the **photosynthetically active radiation** (**PAR**) as it interacts with individual leaves and/or the plant canopy. High spectral resolution imaging spectrometers are useful for measuring the absorption and reflectance characteristics of the photosynthetically active radiation. To demonstrate these principles, consider the spectral reflectance characteristics of four different leaves obtained from a single healthy Sweetgum tree (*Liquidambar styraciflua L.*)

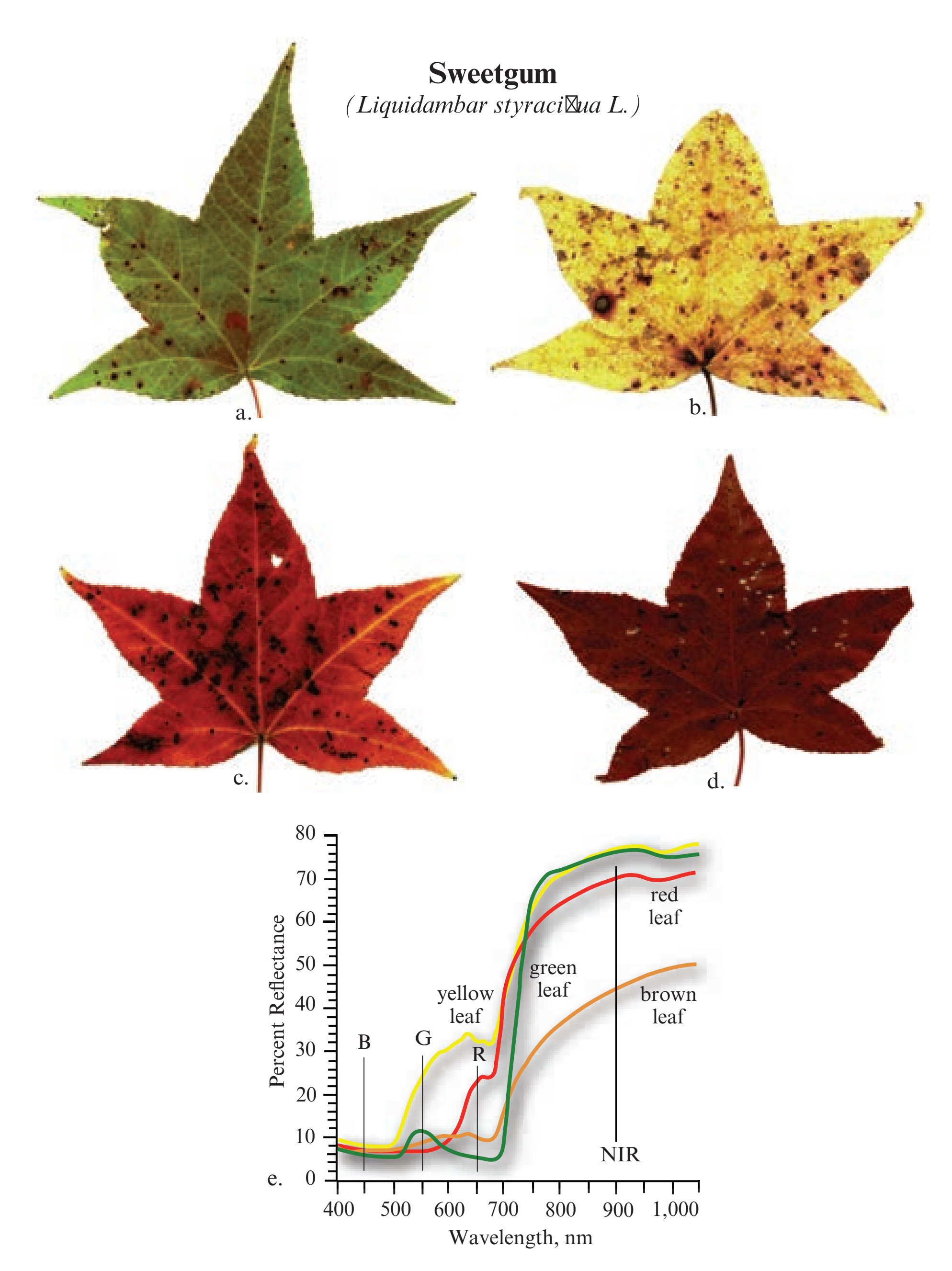


FIGURE 8–37 a) Photosynthesizing green Sweetgum leaf (*Liquidambar styraciflua* L.) obtained from a tree on November 11, 1998. b-c) Senescing yellow and red Sweetgum leaves obtained on the same tree. d) Senesced Sweetgum leaf that was on the ground. e) Spectroradiometer percent reflectance measurements over the wavelength interval 400 – 1,050 nm.

on November 11, 1998, in Columbia, SC (Figure 8-37). The green leaf (a), yellow leaf (b), and red leaf (c) were collected directly from the Sweetgum tree. The dark brown leaf (d) was collected on the ground beneath the tree.

A GER 1500 (Geophysical & Environmental Research, Inc.) handheld spectroradiometer was used to obtain spectral reflectance measurements from each of the leaves. The spectroradiometer obtained spectral reflectance measurements in 512 bands in the ultraviolet, blue, green, red, and near-infrared spectral regions from 350–1,050 nm. Percent reflectance measurements were obtained in the lab by measuring the amount of energy reflected from the surface of the leaf (the target) divided by the amount of energy reflected from a Spectralon® reflectance reference (percent reflectance = target/reference × 100). The reflectance measurements for each of the leaves from 400–1,050 nm were plotted in a percent reflectance graph (Figure 8-37).

The green leaf (Figure 8-37a) was still photosynthesizing and yielded a typical healthy green reflectance spectra with strong chlorophyll absorption bands in

the blue and red regions (approximately 6% reflectance at 450 nm and 5% at 650 nm, respectively), and a peak in reflectance in the green region of the visible spectrum (11% at 550 nm). Approximately 76% of the incident near-infrared radiant flux was reflected from the leaf at 900 nm.

The yellow leaf (Figure 8-37b) was undergoing senescence. As the influence of the chlorophyll pigments diminished, relatively greater amounts of green (24% at 550 nm) and red (32% at 650 nm) light were reflected from the leaf, resulting in a yellow appearance. At 750 nm, the yellow leaf reflected less near-infrared radiant flux than the healthy green leaf. Conversely, the near-infrared reflectance at 900 nm was about 76%, very similar to the healthy green leaf.

The red leaf (Figure 8-37c) reflected 7% of the blue at 450 nm, 6% of the green energy at 550 nm, and 23% of the incident red energy at 650 nm. Near-infrared reflectance at 900 nm dropped to 70%.

The dark brown leaf (Figure 8-37d) produced a spectral reflectance curve with low blue (7% at 450 nm), green (9% at 550 nm), and red reflectance (10% at 650 nm). This combination produced a dark brown appearance. Near-infrared reflectance dropped to 44% at 900 nm.

It is important to understand the physiology of the plants under investigation and especially their pigmentation characteristics so that we can appreciate how a typical plant will appear when chlorophyll absorption starts to decrease, either due to seasonal senescence or environmental stress. As demonstrated, when a plant is under stress and/or chlorophyll production decreases, the lack of chlorophyll pigmentation typically causes the plant to absorb less in the chlorophyll absorption bands. Such plants will have a much higher reflectance, particularly in the green and red portion of the spectrum, and therefore may appear yellowish or *chlorotic*. In fact, Carter (1993) suggests that increased reflectance in the visible spectrum is the most consistent leaf reflectance response to plant stress. Infrared reflectance responds consistently only when stress has developed sufficiently to cause severe leaf dehydration (to be discussed shortly).

Leaf spectral reflectance is most likely to indicate plant stress first in the sensitive 535 – 640 and 685 – 700 nm visible light wavelength ranges. Increased reflectance near 700 nm represents the often reported "blue shift of the red edge," i.e., the shift toward shorter wavelengths of the red-infrared transition curve that occurs in stressed plants when reflectance is plotted versus wavelength (Cibula and Carter, 1992). The shift toward shorter wavelengths in the region from 650 – 700 nm is particularly evident for the yellow and red reflectance curves shown in Figure 8-37e. Remote sensing within these spectrally narrow ranges may provide improved capability to detect plant stress not only in individual leaves but for whole plants and perhaps for densely vegetated canopies (Carter, 1993; Carter et al., 1996).

Normal color film is sensitive to blue, green, and red wavelength energy. Color-infrared film is sensitive to green, red, and near-infrared energy after minus-blue (yellow) filtration. Therefore, even the most simple camera with color or color-infrared film and appropriate band-pass filtration (i.e., only certain wavelengths of light are allowed to pass) can be used to remotely sense differences in spectral reflectance caused by the pigments present in the palisade parenchyma layer of cells in a typical leaf. However, to detect very subtle spectral reflectance differences in the relatively narrow bands suggested by Cibula and Carter (1992) and Carter et al. (1996), it may be necessary to use a high spectral resolution imaging spectroradiometer that has very narrow bandwidths.

Near-Infrared Energy Interaction within the Spongy Mesophyll Cells

In a typical healthy green leaf, the near-infrared reflectance increases dramatically in the region from 700 – 1,200 nm. For example, the healthy green leaf in the previous example reflected approximately 76% of the incident near-infrared energy at 900 nm. Healthy green leaves absorb radiant energy very efficiently in the blue and red portions of the spectrum where incident light is required for photosynthesis. But immediately to the long wavelength side of the red chlorophyll absorption band, why do the reflectance and transmittance of plant leaves increase so dramatically, causing the absorptance to fall to low values (Figure 8-34)? This condition occurs throughout the near-infrared wavelength range where the direct sunlight incident on plants has the bulk of its energy. If plants absorbed this energy with the same efficiency as they do in the visible region, they could become much too warm and the proteins would be irreversibly denatured. As a result, plants have adapted so they do not use this massive amount of near-infrared energy and simply reflect it or transmit it through to underlying leaves or the ground.

The spongy mesophyll layer in a green leaf controls the amount of near-infrared energy that is reflected. The spongy mesophyll layer typically lies below the palisade mesophyll layer and is composed of many cells and intercellular air spaces as shown in Figure 8-35. It is here that the oxygen and carbon dioxide exchange takes place for photosynthesis and respiration. In the near-infrared part of the spectrum, healthy green vegetation is generally characterized by high reflectance (40 – 60%), high transmittance (40 – 60%) through the leaf

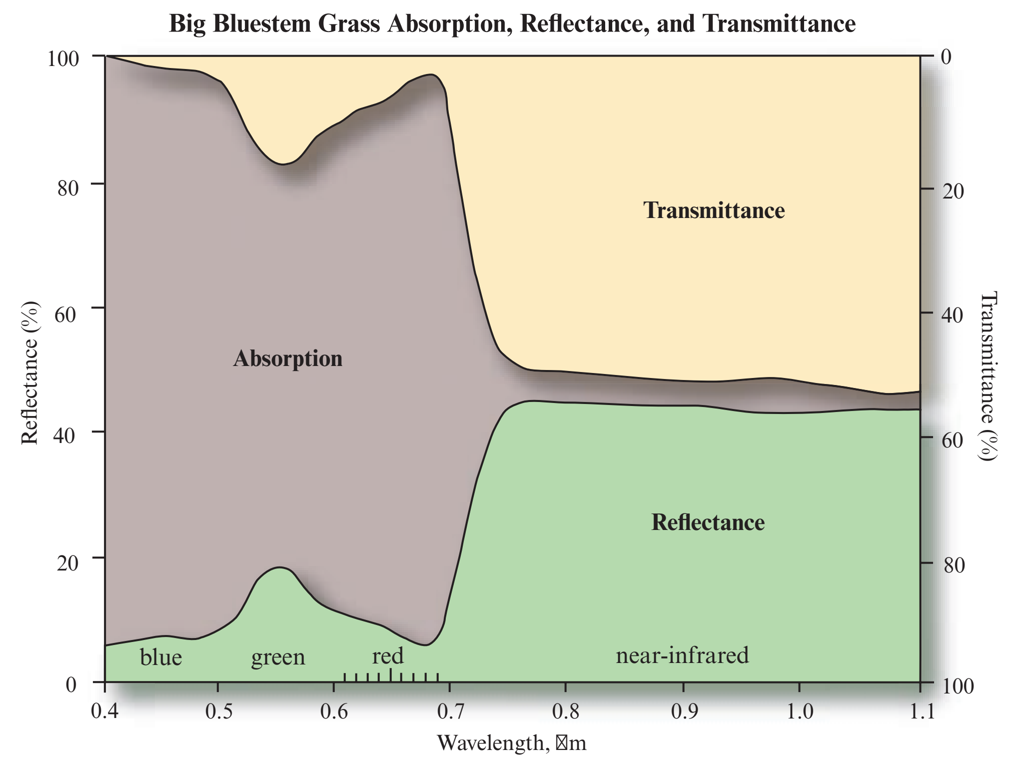


FIGURE 8–38 Hemispheric reflectance, transmittance, and absorption characteristics of Big Bluestem grass adaxial leaf surfaces obtained using a laboratory spectroradiometer. The reflectance and transmittance curves are almost mirror images of one another throughout the visible and near-infrared portions of the electromagnetic spectrum. The blue and red chlorophyll in plants absorbs much of the incident energy in the visible portion of the spectrum (0.4 – 0.7 μm) (based on Walter-Shea and Biehl, 1990). Imaging spectrometers such as AVIRIS are capable of identifying small changes in the absorption and reflection characteristics of plants because the sensors often have channels that are only 10 nm apart, i.e., there could be 10 channels in the region from 0.6 to 0.7 μm (600 – 700 nm).

onto underlying leaves, and relatively low absorptance (5–10%). Notice that a healthy green leaf's reflectance and transmittance spectra throughout the visible and near-infrared portion of the spectrum are almost mirror images of one another, as shown in Figure 8-38 (Walter-Shea and Biehl, 1990).

The high diffuse reflectance of the near-infrared (0.7 – 1.2 μm) energy from plant leaves is due to the internal scattering at the cell wall–air interfaces within the leaf (Gausmann et al., 1969; Peterson and Running, 1989). A well-known water vapor absorption band exists at 0.92 – 0.98 μm; consequently, the optimum spectral region for sensing in the near-infrared region is believed to be 0.74 – 0.90 μm (Tucker, 1979).

The main reasons that healthy plant canopies reflect so much near-infrared energy are:

- the leaf already reflects 40–60% of the incident near-infrared energy from the spongy mesophyll (Figure 8-34), and
- the remaining 45–50% of the energy penetrates (i.e., is transmitted) through the leaf and can be reflected once again by leaves below it.

This is called *leaf additive reflectance*. For example, consider the reflectance and transmission characteristics of the hypothetical two-layer plant canopy shown in Figure 8-39. Assume that leaf 1 reflects 50% of the incident near-infrared energy back into the atmosphere and that the remaining 50% of the near-infrared energy is transmitted through leaf 1 onto leaf 2. The transmitted energy then falls on leaf 2 where 50% again is transmitted (25% of the original) and 50% is reflected. The reflected energy then passes back through leaf 1, which allows half of that energy (or 12.5% of the original) to be transmitted and half reflected. The resulting total energy exiting leaf 1 in this two-layer example is 62.5% of the incident energy. Therefore, the greater the number of leaf layers in a healthy, mature canopy, theoretically the greater the infrared reflectance. Conversely, if the canopy is only composed of a single, sparse leaf layer then the near-infrared reflectance will not be as great because the energy that is transmitted through the leaf layer may be absorbed by the ground cover beneath the canopy.

Changes in the near-infrared spectral properties of healthy green vegetation may provide information about plant senescence and/or stress. For example,

FIGURE 8–39 A hypothetical example of additive reflectance from a canopy with two leaf layers. Fifty percent of the incident radiant flux, Φ_i, to leaf 1 is reflected (R_1) and the other 50% is transmitted onto leaf 2 (T_1). Fifty percent of the radiant flux incident to leaf 2 is transmitted through leaf 2 (T_2); the other 50% is reflected toward the base of leaf 1 (R_2). Fifty percent of the energy incident at the base of leaf 1 is transmitted through it (T_3) while the remaining 50% (R_3) is reflected toward leaf 2 once again. At this point, an additional 12.5% (1/8) reflectance has been contributed by leaf 2, bringing the total reflected radiant flux to 62.5%. However, to be even more accurate, one would have to also take into account the amount of energy reflected from the base of leaf 1 (R_3) onto leaf 2, and the amount reflected from leaf 2 (R_4) and eventually transmitted through leaf 1 once again (T_5). This process would continue.

consider the four leaves and their spectral reflectance characteristics shown in Figure 8-37. The photosynthesizing green leaf (a) exhibited strong chlorophyll absorption in the blue and red wavelength regions, an understandable increase in green reflectance, and approximately 76% reflectance in the near-infrared region. After a certain point, near-infrared reflectance decreased as the leaves senesced (b–d). However, if the leaves were to dry out significantly during senescence, we would expect to see much higher reflectance in the near-infrared region (to be discussed shortly).

Scientists have known since the 1960s that a direct relationship exists between response in the near-infrared region and various biomass measurements. Conversely, it has been shown that an inverse relationship exists between the response in the visible region, particularly the red, and plant biomass. The best way to appreciate this is to plot all of the pixels in a typical remote sensing scene in red and near-infrared reflectance space. For example, Figure 8-40a suggests where approximately 10,000 pixels in a typical agricultural scene are located in red and near-infrared multispectral feature space (i.e., within the green area). Dry bare soil fields and wet bare soil fields in the scene would be located at opposite ends of the soil line. This means that a wet bare soil would have very low red and near-infrared reflectance. Conversely, a dry bare soil area would probably have high red and high near-infrared reflectance. As a vegetation canopy matures, it reflects more near-infrared energy while at the same time absorbing more red radiant flux for photosynthetic purposes. This causes the spectral reflectance of the pixel to move in a perpendicular direction away from the soil line. As biomass increases and as the plant canopy cover increases, the field's location in the red and near-infrared spectral space moves farther away from the soil line.

Figure 8-40b demonstrates how just one agricultural pixel might move about in the red and near-infrared spectral space during a typical growing season. If the field was prepared properly, it would probably be located in the moist bare soil region of the soil line with low red and low near-infrared reflectance at the beginning

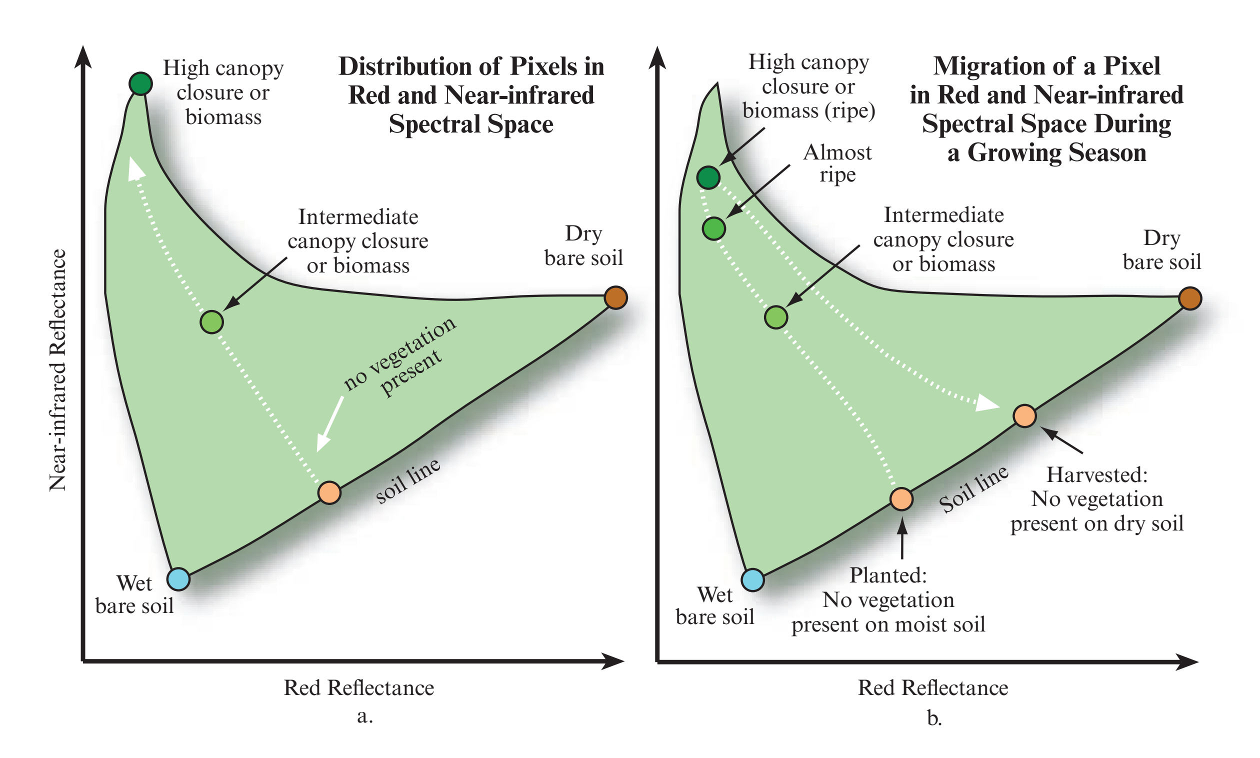


FIGURE 8–40 a) The distribution of all the pixels in a scene in red and near-infrared multispectral feature space is found in the green shaded area. Wet and moist bare-soil fields are located along the soil line. The greater the biomass and/or crop canopy closure, the greater the near-infrared reflectance and the lower the red reflectance. This condition moves the pixel's spectral location a perpendicular direction away from the soil line. b) The migration of a single vegetated agricultural pixel in red and near-infrared multispectral feature space during a growing season is shown. After the crop emerges, it departs from the soil line, eventually reaching complete canopy closure. After harvesting, the pixel will be found on the soil line, but perhaps in a drier soil condition.

of the growing season. After the crop emerges from the ground, it would depart from the soil line, eventually reaching complete canopy closure. At this point, the reflected near-infrared radiant flux would be high and the red reflectance would be low. After harvesting, the pixel would probably be found once again on the soil line but perhaps in a drier condition.

The relationship between red and near-infrared canopy reflectance has resulted in the development of remote sensing vegetation indices and biomass-estimating techniques that utilize multiple measurements in the visible and near-infrared region (e.g., Liu et al, 2012). The result is a linear combination that may be more highly correlated with biomass than either red or near-infrared measurement alone.

Middle-Infrared Energy Interaction with Water in the Spongy Mesophyll

Plants require water to grow. A leaf obtains water through the plant's roots. The water travels from the roots, up the stem, and enters the leaf through the *petiole*. Veins carry water to the cells in the leaf. If a plant is watered so that it contains as much water as it can possibly hold at a given time, it is said to be fully *turgid*. Much of the water is found in the spongy mesophyll portion of the plant. If we forget to water the plant or rainfall decreases, the plant will contain an amount of water that is less than it can potentially hold. This is called its relative *turgidity*. It would be useful to have a remote sensing instrument that was sensitive to how much water was actually in a plant leaf. Remote sensing in the middle-infrared, thermal infrared, and passive microwave portion of the electromagnetic spectrum can provide such information to a limited extent.

Water in the atmosphere creates five major absorption bands in the near-infrared through middle-infrared portions of the electromagnetic spectrum at 0.97, 1.19, 1.45, 1.94, and 2.7 μm (Figure 8-34). The fundamental vibrational water-absorption band at 2.7 μm is the strongest in this part of the spectrum (there is also one in the thermal infrared region at 6.27 μm). However, there is also a strong relationship between the reflectance in the middle-infrared region from 1.3 – 2.5 μm and the amount of water present in the leaves of a plant canopy. Water in plants absorbs incident energy between the absorption bands with increasing strength at longer wavelengths. In these middle-infrared wavelengths, vegetation reflectance peaks occur at about 1.6

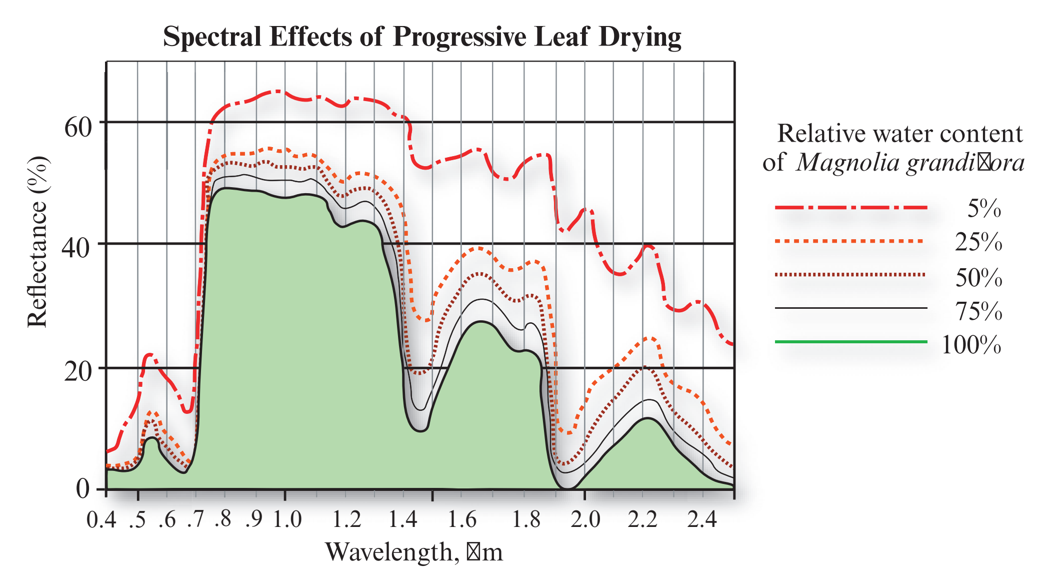


FIGURE 8–41 Reflectance response of a single magnolia leaf (*Magnolia grandiflora*) to decreased relative water content. As moisture content decreased, reflectance increased throughout the 0.4 to 2.5 μm region. However, the greatest increase occurred in the middle-infrared region from 1.3 to 2.5 μm (based on Carter, 1991).

and 2.2 μm, between the major atmospheric water absorption bands (Figure 8-34).

Water is a very good absorber of middle-infrared energy, so generally, the greater the turgidity of the leaves, the lower the middle-infrared reflectance. Conversely, as the moisture content of leaves decreases, reflectance in the middle-infrared region increases substantially. As the amount of plant water in the intercellular air spaces decreases, this causes the incident middle-infrared energy to be more intensely scattered at the interface of the intercellular walls resulting in much greater middle-infrared reflectance from the leaf. For example, consider the spectral reflectance of Magnolia leaf samples at five different moisture conditions displayed over the wavelength interval from 0.4 – 2.5 μm (Figure 8-41). The middle-infrared wavelength intervals from about 1.5 – 1.8 μm and from 2.1 – 2.3 μm appear to be more sensitive to changes in the moisture content of the plants than the visible or near-infrared portions of the spectrum (i.e., the *y*-axis distance between the spectral reflectance curves is greater as the moisture content decreases). Also note that substantive changes in the visible reflectance curves (0.4 – 0.7 μm) did not begin to appear until the plant moisture in the leaves decreased to about 50%. When the relative water content of the plant decreases to approximately 50%, almost any portion of the visible, near- and middle-infrared regions might provide some valuable spectral reflectance information.

Reflectance in the middle-infrared region is inversely related to the absorptance of a layer of water approximately 1 mm in depth (Carter, 1991). The degree to which incident solar energy in the middle-infrared region is absorbed by vegetation is a function of the total amount of water present in the leaf and the leaf thickness. If proper choices of sensors and spectral bands are made, it is possible to monitor the relative turgidity in plants.

Most optical remote sensing systems (except radar) are generally constrained to function in the wavelength intervals from 0.3 – 1.3, 1.5 – 1.8, and 2.0 – 2.6 μm due to the strong atmospheric water absorption bands at 1.45, 1.94, and 2.7 μm. Fortunately, as demonstrated in Figure 8-34 there is a strong "carryover" sensitivity to water content in the 1.5 – 1.8 and 2.0 – 2.6 μm regions adjacent to the major water absorption bands. This is the main reason that the Landsat Thematic Mapper (4 and 5) and Landsat 7 Enhanced Thematic Mapper Plus (ETM^+) were made sensitive to two bands in this region: band 5 (1.55 – 1.75 μm) and band 7 (2.08 – 2.35 μm). The 1.55 – 1.75 μm middle-infrared band has consistently demonstrated a sensitivity to canopy moisture content. For example, Pierce et al. (1990) found that this band and vegetation indices produced using it were correlated with canopy water stress in coniferous forests.

Much of the water in a plant is lost via transpiration. *Transpiration* occurs as the Sun warms the water inside the leaf, causing some of the water to change its state to water vapor that escapes through the stomata. The following are several important functions that transpiration performs:

- It cools the inside of the leaf because the escaping water vapor contains heat.

- It keeps water flowing up from the roots, through the stem, to the leaves.
- It ensures a steady supply of dissolved minerals from the soil.

As molecules of water vapor at the top of the leaf in the tree are lost to transpiration, the entire column of water is pulled upward. Plants lose a considerable amount of water through transpiration each day. For example, a single corn plant can lose about 4 quarts (3.8 liters) of water on a very hot day. If the roots of the plant cannot replace this water, the leaves wilt, photosynthesis stops, and the plant dies. Thus, monitoring the moisture content of plant canopies, which is correlated with rates of transpiration, can provide valuable information on the health of a crop or stand of vegetation. Thermal infrared and passive microwave remote sensing have provided valuable plant canopy evapotranspiration information.

The most practical application of plant moisture information is the regional assessment of crop water conditions for irrigation scheduling, stress assessment, and yield modeling for agriculture, rangeland, and forestry management.

Remote Sensing–Derived Vegetation Indices

Since the 1960s, scientists have extracted and modeled various vegetation biophysical variables using remotely sensed data. Much of this effort has involved the use of **vegetation indices**—dimensionless, radiometric measures that indicate the relative abundance and activity of green vegetation, including leaf-area-index (LAI), percentage green cover, chlorophyll content, green biomass, and absorbed photosynthetically active radiation (APAR). A vegetation index should (Running et al., 1994; Huete and Justice, 1999):

- maximize sensitivity to plant biophysical parameters, preferably with a linear response in order that sensitivity be available for a wide range of vegetation conditions, and to facilitate validation and calibration of the index;
- normalize or model external effects such as Sun angle, viewing angle, and the atmosphere for consistent spatial and temporal comparisons;
- normalize internal effects such as canopy background variations, including topography (slope and aspect), soil variations, and differences in senesced or woody vegetation (nonphotosynthetic canopy components);
- be coupled to some specific measurable biophysical parameter such as biomass, LAI, or APAR as part of the validation effort and quality control.

There are many vegetation indices. Good summaries about vegetation indices are found in Cheng et al. (2008), Galvao et al., (2009), Vina et al. (2011), Gray and Song (2012), Liu et al. (2012), and Wang et al. (2013). Many are functionally equivalent (redundant) in information content (Perry and Lautenschlager, 1984), and some provide unique biophysical information. It is useful to review the historical development of some of the major indices and provide information about recent advances in index development. A few of the most widely adopted are summarized in Table 8-9.

Simple Ratio—SR

Many indices make use of the inverse relationship between red and near-infrared reflectance associated with healthy green vegetation (Figure 8-42a). Cohen (1991) suggests that the first true vegetation index was the *Simple Ratio* (SR), which is the ratio of red reflected radiant flux (ρ_{red}) to near-infrared radiant flux (ρ_{nir}) as described in Birth and McVey (1968):

$$SR = \frac{\rho_{red}}{\rho_{nir}}. \tag{8.36}$$

The simple ratio provides valuable information about vegetation biomass or LAI (Schlerf et al., 2005; Pena-Barragan et al., 2011; Gray and Song, 2012). It is especially sensitive to biomass and/or LAI variations in high-biomass vegetation such as forests (Huete et al., 2002b).

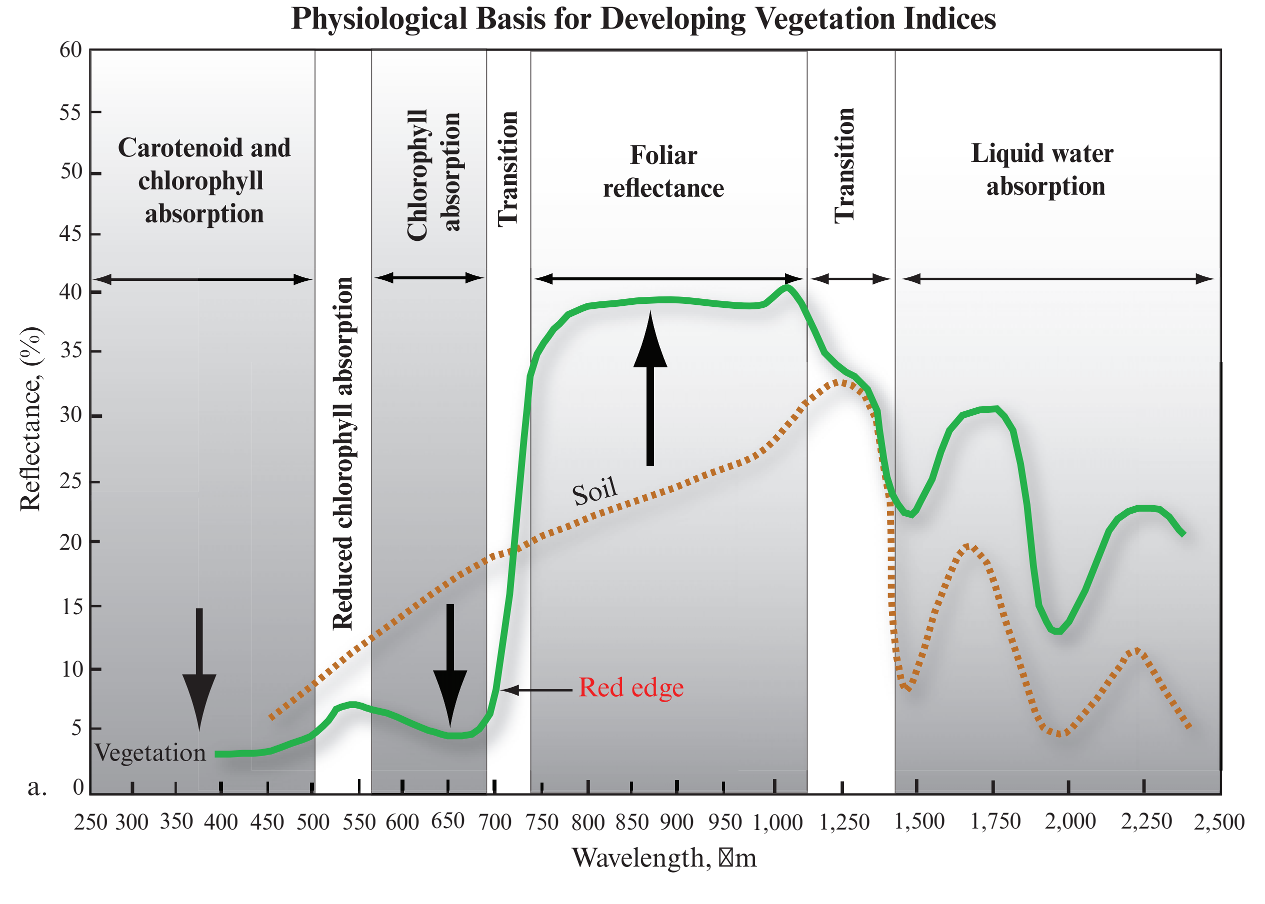
Normalized Difference Vegetation Index—NDVI

Rouse et al. (1974) were some of the first to use what has become known as the *Normalized Difference Vegetation Index* (NDVI):

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}. \tag{8.37}$$

The NDVI is functionally equivalent to the simple ratio; that is, there is no scatter in an SR vs. NDVI plot, and each SR value has a fixed NDVI value. When we plot the mean NDVI and SR values for various biomes, we find that the NDVI approximates a nonlinear transform of the simple ratio (Figure 8-42b) (Huete et al., 2002b). The NDVI is an important vegetation index because:

- Seasonal and inter-annual changes in vegetation growth and activity can be monitored.
- The ratioing reduces many forms of *multiplicative* noise (Sun illumination differences, cloud shadows, some atmospheric attenuation, some topographic variations) present in multiple bands of multiple-date imagery.



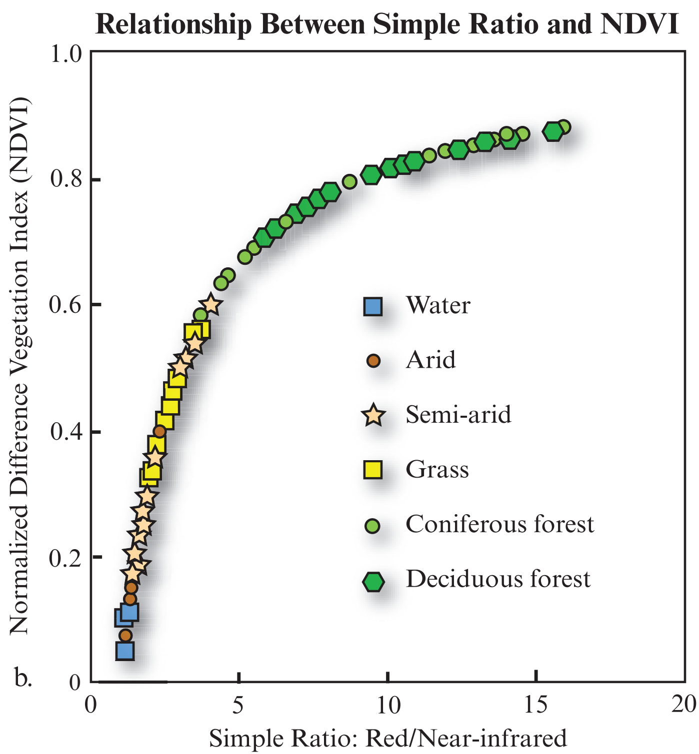


FIGURE 8–42 a) The physiological basis for developing vegetation indices. Typical spectral reflectance characteristics for healthy green grass and bare dry soil for the wavelength interval from 250 – 2,500 nm. b) The NDVI is a normalized ratio of the near-infrared and red bands. The NDVI is functionally equi valent to and is a nonl inear transform of the simple ratio (based on Huete et al., 2002b).

Following are disadvantages of the NDVI (Huete et al., 2002a; Wang et al., 2004):

- The ratio-based index is nonlinear and can be influenced by *additive* noise effects such as atmospheric path radiance (Chapter 6).
- NDVI is usually highly correlated with vegetation leaf-area-index (LAI). However, the relationship may not be as strong during periods of maximum LAI, apparently due to the saturation of NDVI when LAI is very high (Wang et al., 2004). For example, Figure 8-42b reveals that the NDVI dynamic range is stretched in favor of low-biomass conditions and compressed in high-biomass, forested regions. The opposite is true for the Simple Ratio, in which most of the dynamic range encompasses the high-biomass forests with little variation reserved for the lower-biomass regions (grassland, semi-arid, and arid biomes).
- It is very sensitive to canopy background variations (e.g., soil visible through the canopy). NDVI values are particularly high when darker-canopy backgrounds exist.

Scientists continue to use the NDVI heavily (e.g. Galvao et al., 2005, 2009; Neigh et al., 2008; Mand et al., 2010; Sonnenschein et al., 2011; Vina et al., 2011, Liu et al., 2012; Li et al., 2014). In fact, two of the standard MODIS land products are sixteen-day composite NDVI datasets of the world at a spatial resolution of 500 m and 1 km (Huete et al., 2002a). NDVI images derived from MODIS data obtained in February and October, 2012, are shown in Figure 8-43. Note the "greening-up" of the terrain in the Northern hemisphere, especially in Canada, Europe, and Russia. Also note the greening-up of the Sahel region of Africa just below the Sahara Desert. Geospatial biophysical information such as this is used in many biophysical models.

The NDVI transform also performs well on high spatial resolution imagery that has red and near-infrared bands. For example, consider the NDVI image extracted from the color-infrared 1 × 1 ft. spatial resolution digital frame camera photography shown in Figure 8-44. The single-band NDVI image is displayed in Figure 8-44e. The brighter the shade of gray in the NDVI image, the greater the amount of green biomass. Note that the water in the swimming pools, the street concrete, and building rooftops all have low NDVI values (i.e., they are dark). High NDVI values associated with vegetation are very easy to locate in the color density-sliced image shown in Figure 8-44f.

Kauth-Thomas Tasseled Cap Transformation

Kauth and Thomas (1976) developed an orthogonal transformation of the original Landsat MSS data space to a new four-dimensional feature space. It was called the *tasseled cap* or *Kauth-Thomas transformation*. It created four new axes: the soil brightness index (*B*), greenness vegetation index (*G*), yellow stuff index (*Y*), and none-such (*N*). The names attached to the new axes indicate the characteristics the indices were intended to measure. The coefficients are (Kauth et al., 1979):

$$B = 0.332MSS1 + 0.603MSS2 + 0.675MSS3 + 0.262MSS4 \quad (8.38)$$

$$G = -0.283MSS1 - 0.660MSS2 + 0.577MSS3 + 0.388MSS4 \quad (8.39)$$

$$Y = -0.899MSS1 + 0.428MSS2 + 0.076MSS3 - 0.041MSS4 \quad (8.40)$$

$$N = -0.016MSS1 + 0.131MSS2 - 0.452MSS3 + 0.882MSS4\,. \quad (8.41)$$

Crist and Kauth (1986) determined the visible, near-infrared, and middle-infrared coefficients for transforming Landsat Thematic Mapper imagery into brightness, greenness, and wetness variables:

$$B = 0.2909TM1 + 0.2493TM2 + 0.4806TM3 + 0.5568TM4 + 0.4438TM5 + 0.1706TM7 \quad (8.42)$$

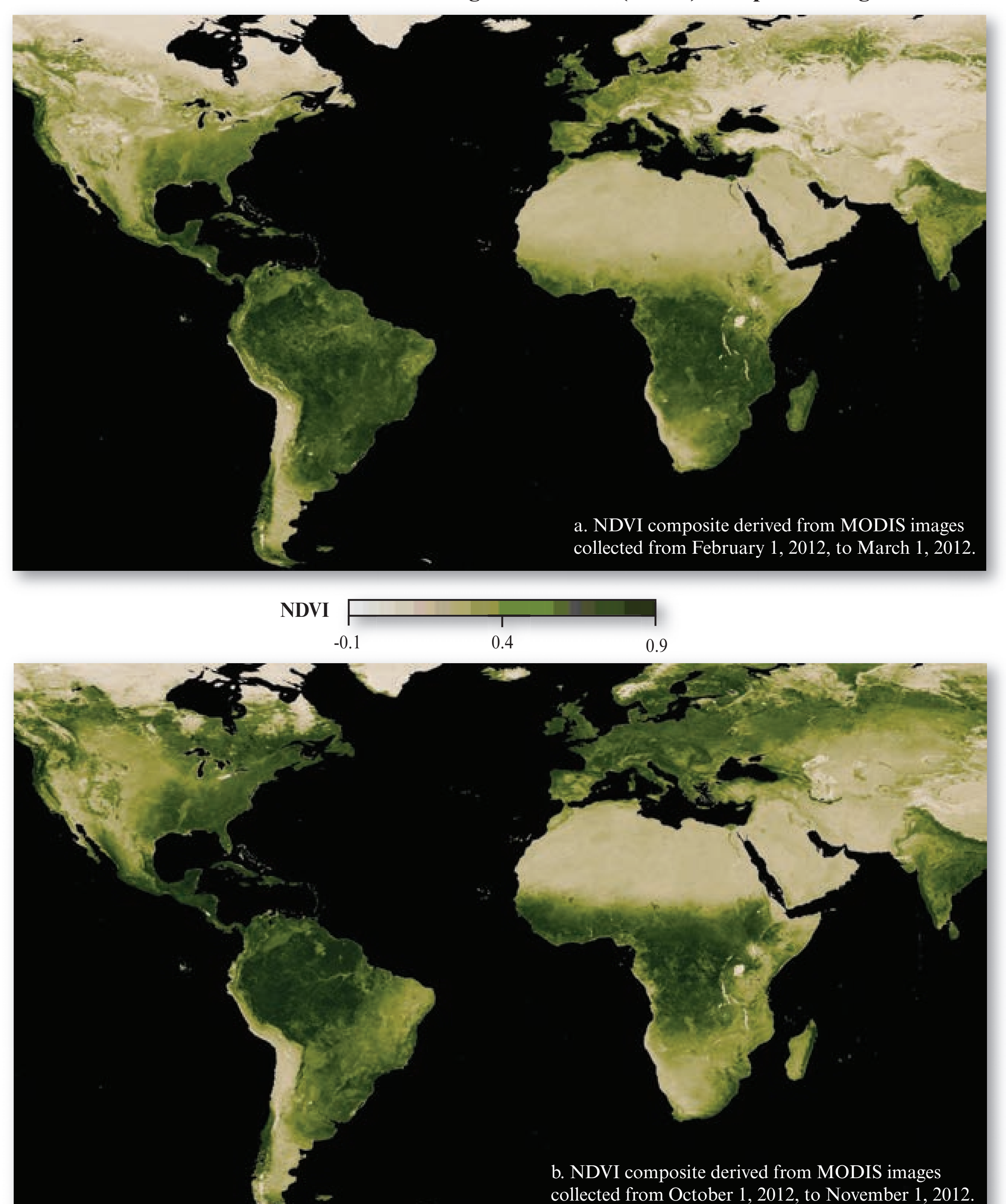
$$G = -0.2728TM1 - 0.2174TM2 - 0.5508TM3 + 0.7221TM4 + 0.0733TM5 - 0.1648TM7 \quad (8.43)$$

$$W = 0.1446TM1 + 0.1761TM2 + 0.3322TM3 + 0.3396TM4 - 0.6210TM5 - 0.4186TM7\,. \quad (8.44)$$

It is called the tasseled cap transformation because of its unique cap shape (Yarbrough et al., 2012) (Figures 8-45a–c). Crist and Cicone (1984) identified a third component that is related to soil features, including moisture status (Figure 8-45d). Thus, an important source of soil information is available through the inclusion of the middle-infrared bands of the Thematic Mapper (Price et al., 2002).

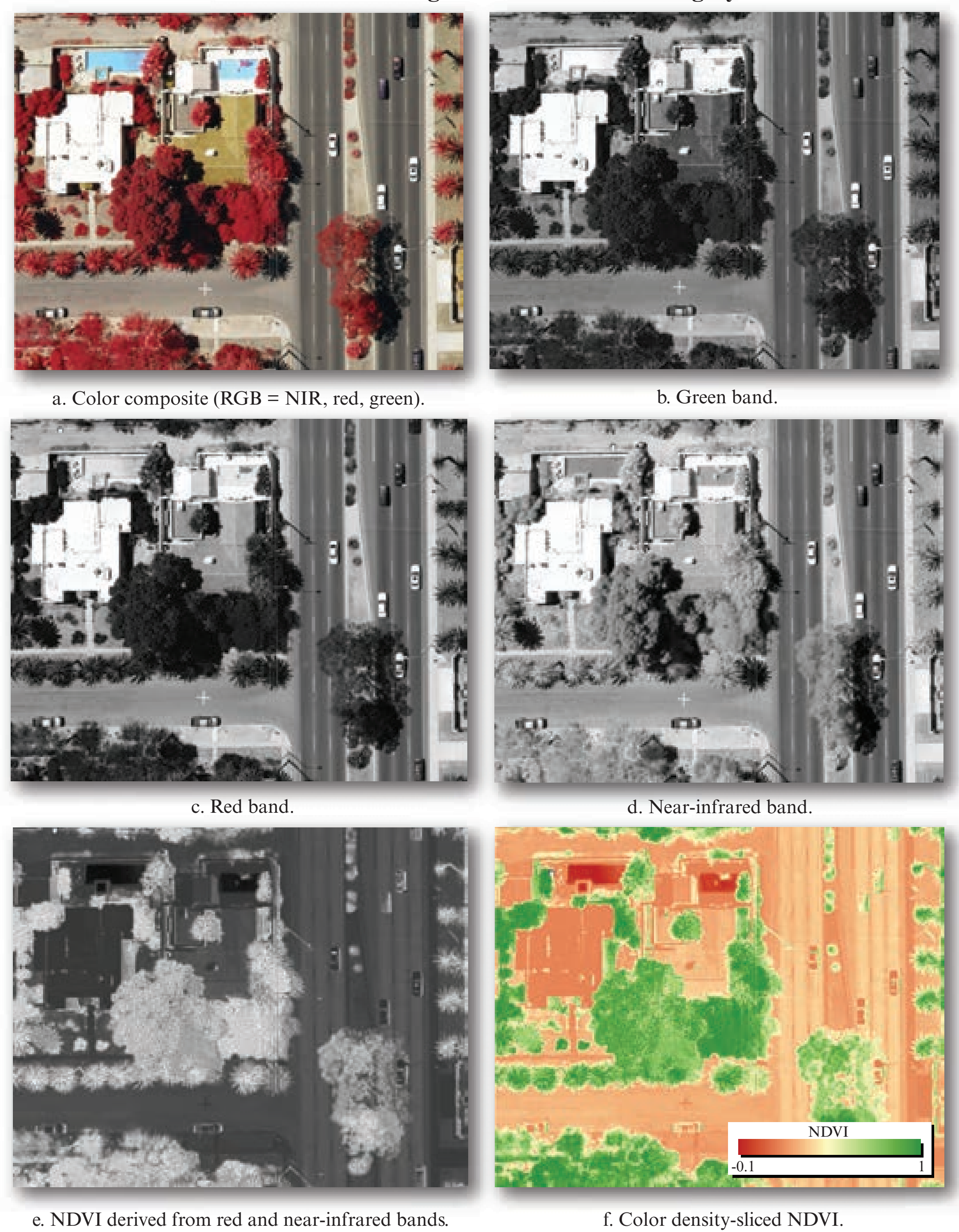
The 1982 Charleston, SC, Landsat TM scene was transformed into brightness, greenness, and wetness (moisture) content images (Figure 8-46) based on the use of the TM tasseled cap coefficients (Equations 8-42 through 8-44). Urbanized areas are particularly evident in the brightness image. The greater the biomass, the brighter the pixel value in the greenness image. The wetness image provides subtle information concerning the moisture status of the wetland environment. As expected, the greater the moisture content, the brighter the response. A color composite of the dataset is shown in Figure 8-46d (RGB = brightness, greenness, and wetness).

The tasseled cap transformation is a global vegetation index. Theoretically, it can be used anywhere in the world to disaggregate the amount of soil brightness, vegetation, and moisture content in individual pixels in Landsat MSS, TM, ETM^{+}, Landsat 8 or other types of



FIGU RE 8–43 a) NDVI composite image derived from MODIS data acquired from February 1, 2012, to March 1, 2012. The darker the green, the higher the NDVI value and the greater the leaf-area-index (LAI) and biomass. b). NDVI composite image derived from MODIS data acquired from October 1, 2012, to November 1, 2012 (imagery courtesy of NASA Earth Observations - NEO).

Normalized Difference Vegetation Index (NDVI) Image Extracted from UltraCam® Digital Frame Camera Imagery



a. Color composite (RGB = NIR, red, green).

b. Green band.

c. Red band.

d. Near-infrared band.

e. NDVI derived from red and near-infrared bands.

f. Color density-sliced NDVI.

FIGURE 8–44 An example of the NDVI transformation applied to color-infrared 1 × 1 ft. spatial resolution digital frame camera imagery (original aerial photography courtesy of John Copple, Sanborn Map Company). a) Original image. b–d). Green, red, and near-infrared bands. e) NDVI image. f) Density-sliced NDVI image.

multispectral data. Practically, however, it is better to compute the coefficients based on local conditions if possible. Jackson (1983) provided a computer program for this purpose.

The Kauth-Thomas tasseled cap transformation continues to be widely used (e.g., Powell et al., 2010; Sonnenschein et al., 2011; Pflugmacher et al., 2012; Yarbrough et al., 2012, 2014; Ali Baig et al., 2014). Huang et al. (2002) developed tasseled cap coefficients for use with Landsat 7 ETM$^+$ at-satellite reflectance data (Table 8-10). These coefficients are most appropriate for regional applications where atmospheric correction is not feasible. Yarbrough et al. (2012) developed a

TABLE 8–9 Selected remote sensing vegetation and other indices.

Vegetation Index	Equation	Selected References
Simple Ratio (SR)	$SR = \frac{\rho_{red}}{\rho_{nir}}$	Birth and McVey, 1968 Colombo et al., 2003 Schlerf et al., 2005 Boer et al., 2008 Pena-Barragan, 2011 Vina et al., 2011 Gray and Song, 2012
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}$	Rouse et al., 1974 Deering et al., 1975 Huete et al., 2002a Colombo et al., 2003 Schlerf et al., 2005 Houborg et al., 2007 Boer et al., 2008 Cheng et al., 2008 Neigh et al., 2008 Mand et al., 2010 Sonnenschein et al., 2011 Pena-Barragan, 2011 Purkis and Klemas, 2011 Sims et al., 2011 Vina et al., 2011 Gray and Song, 2012 Liu et al., 2012 Li et al., 2014
Kauth-Thomas Transformation Brightness Greenness Yellow stuff *None-such*	Landsat Multispectral Scanner (MSS) $B = 0.332 MSS1 + 0.603 MSS2 + 0.675 MSS3 + 0.262 MSS4$ $G = -0.283 MSS1 - 0.660 MSS2 + 0.577 MSS3 + 0.388 MSS4$ $Y = -0.899 MSS1 + 0.428 MSS2 + 0.076 MSS3 - 0.041 MSS4$ $N = -0.016 MSS1 + 0.131 MSS2 - 0.452 MSS3 + 0.882 MSS4$	Kauth and Thomas, 1976 Kauth et al., 1979
Brightness Greenness Wetness	Landsat Thematic Mapper (TM) $B = 0.2909 TM1 + 0.2493 TM2 + 0.4806 TM3 + 0.5568 TM4 + 0.4438 TM5 + 0.1706 TM7$ $G = -0.2728 TM1 - 0.2174 TM2 - 0.5508 TM3 + 0.7221 TM4 + 0.0733 TM5 - 0.1648 TM7$ $W = 0.1446 TM1 + 0.1761 TM2 + 0.3322 TM3 + 0.3396 TM4 - 0.6210 TM5 - 0.4186 TM7$	Crist and Kauth, 1986 Price et al., 2002 Rogan et al., 2002 Jin and Sader, 2005 Schlerf et al., 2005 Powell et al., 2010 Sonnenschein et al., 2011 Pflugmacher et al., 2012 Yarbrough et al., 2012 Ali Baig et al., 2014 Yarbrough, 2014
Normalized Difference Moisture or Water Index (NDMI or NDWI)	$NDMI \text{ or } NDWI = \frac{(\rho_{nir} - \rho_{swir})}{(\rho_{nir} + \rho_{swir})}$	Hardisky et al., 1983 Gao, 1996; Jackson et al., 2004 Galvao et al., 2005 Jin and Sader, 2005 Houborg et al., 2007 Verbesselt et al., 2007 Cheng et al., 2008 Gray and Song, 2012 Wang et al., 2013
Perpendicular Vegetation Index (PVI)	$PVI = \sqrt{(0.355 MSS4 - 0.149 MSS2)^2 + (0.355 MSS2 - 0.852 MSS4)^2}$	Richardson and Wiegand, 1977
	$PVI = \frac{(NIR - aRed - b)}{\sqrt{1 + a^2}}$	Colombo et al., 2003 Guyon et al., 2011
Leaf Relative Water Content Index (LWCI)	$LWCI = \frac{-\log[1 - (NIR_{TM4} - MidIR_{TM5})]}{-\log[1 - (NIR_{TM4_{ft}} - MidIR_{TM5_{ft}})]}$	Hunt et al., 1987

TABLE 8–9 Selected remote sensing vegetation and other indices.

Vegetation Index	Equation	Selected References
Soil Adjusted Vegetation Index (SAVI) and Modified SAVI (MSAVI)	$SAVI = \frac{(\rho_{nir} - \rho_{red})}{(\rho_{nir} + \rho_{red} + L)}(1 + L)$	Huete, 1988 Huete and Liu, 1994 Running et al., 1994 Qi et al., 1995 Colombo et al., 2003 Sonnenschein et al., 2011
Atmospherically Resistant Vegetation Index (ARVI)	$ARVI = \left(\frac{\rho_{nir} - \rho_{rb}}{\rho_{nir} + \rho_{rb}}\right)$	Kaufman and Tanre, 1992 Huete and Liu, 1994 Colombo et al., 2003
Soil and Atmospherically Resistant Vegetation Index (SARVI)	$SARVI = \frac{\rho_{nir} - \rho_{rb}}{\rho_{nir} + \rho_{rb} + L}$	Huete and Liu, 1994; Running et al., 1994
Enhanced Vegetation Index (EVI)	$EVI = G\frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + C_1\rho_{red} - C_2\rho_{blue} + L}$ $EVI2 = 2.5\frac{(\rho_{nir} - \rho_{red})}{(\rho_{nir} + 2.4\rho_{red} + 1)}$	Huete et al., 1997 Huete and Justice, 1999 Huete et al., 2002a Colombo et al., 2003 TBRS, 2003 Houborg et al., 2007 Wardlow et al., 2007 Cheng, 2008 Jiang et al., 2008 Sims et al., 2011 Gray and Song, 2012 Liu et al., 2012 Wagle et al., 2014
Aerosol Free Vegetation Index (AFRI)	$AFRI_{1.6\mu m} = \frac{(\rho_{nir} - 0.66\rho_{1.6\mu m})}{(\rho_{nir} + 0.66\rho_{1.6\mu m})}$ $AFRI_{2.1\mu m} = \frac{(\rho_{nir} - 0.5\rho_{2.1\mu m})}{(\rho_{nir} + 0.5\rho_{2.1\mu m})}$	Karnieli et al., 2001
Triangular Vegetation Index (TVI)	$TVI = \frac{1}{2}(120(\rho_{nir} - \rho_{green})) - 200(\rho_{red} - \rho_{green})$	Broge and Leblanc, 2000 Pena-Barragan et al., 2011
Reduced Simple Ratio (RSR)	$RSR = \frac{\rho_{nir}}{\rho_{red}}\left(1 - \frac{\rho_{swir} - \rho_{swirmin}}{\rho_{swirmax} + \rho_{swirmin}}\right)$	Chen et al., 2002 Gray and Song, 2012
Ratio TCARI /OSAVI	$TCARI = 3\left[(\rho_{700} - \rho_{670}) - 0.2(\rho_{700} - \rho_{550})\left(\frac{\rho_{700}}{\rho_{670}}\right)\right]$ $OSAVI = \frac{(1 + 0.16)(\rho_{800} - \rho_{670})}{(\rho_{800} + \rho_{670} + 0.16)}$ $\frac{TCARI}{OSAVI}$	Kim et al., 1994 Rondeaux et al., 1996 Daughtry et al., 2000 Haboudane et al., 2002 Pena-Barragan, 2011
Visible Atmospherically Resistant Index (VARI)	$VARI_{green} = \frac{\rho_{green} - \rho_{red}}{\rho_{green} + \rho_{red} - \rho_{blue}}$	Gitelson et al., 2002

TABLE 8–9 Selected remote sensing vegetation and other indices.

Vegetation Index	Equation	Selected References
Normalized Difference Built-up Index (NDBI)	$NDBI = \frac{MidIR_{TM5} - NIR_{TM4}}{MidIR_{TM5} + NIR_{TM4}}$ $\text{built-up}_{area} = NDBI - NDVI$	Zha et al., 2003
Vegetation Adjusted Nighttime Lights (NTL) Urban Index (VANUI)	$VANUI = (1 - NDVI) \times NTL$	Zhang et al., 2013
Red-edge Position	$REP = 700 + 40\left[\frac{\rho_{(\text{red edge})} - \rho_{(700\text{nm})}}{\rho_{(740\text{nm})} - \rho_{(700\text{nm})}}\right]$ where $\rho_{(\text{red edge})} = \frac{\rho_{(670\text{nm})} + \rho_{(780\text{nm})}}{2}$	Clevers, 1994 Dawson and Curran, 1998 Baranoski, 2005
Photochemical Reflectance Index (PRI)	$PRI = \frac{(\rho_{531} - \rho_{570})}{(\rho_{531} + \rho_{570})}$	Mand et al. 2010 Sims et al., 2011
NDVI and Cellulose Absorption Index (CAI) to Quantify Fractional Cover of Photosynthetic Vegetation, Non-photosynthetic Vegetation and Bare Soil	$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}$ and $CAI = [0.5 \times (\rho_{2.0} + \rho_{2.2}) - \rho_{2.1}] \times 10$	Guerschman et al., 2009 Wang et al., 2013
MERIS Terrestrial Chlorophyll Index (MTCI)	$MTCI = \frac{\rho_{band\ 10} - \rho_{band\ 9}}{\rho_{band\ 9} - \rho_{band\ 8}} = \frac{\rho_{753.75} - \rho_{708.75}}{\rho_{708.75} - \rho_{681.25}}$	Dash et al., 2010
Normalized Burn Ratio (NBR)	$NBR = \frac{\rho_{nir} - \rho_{swir}}{\rho_{nir} + \rho_{swir}}$ $\Delta NBR = NBR_{pre-fire} - NBR_{post-fire}$	Brewer et al., 2005 Boer et al., 2008

Kauth-Thomas transform workflow that includes image screening methods to determine if scenes are appropriate for application of the transformation. ERDAS Imagine and ENVI software provide Kauth-Thomas transformation algorithms.

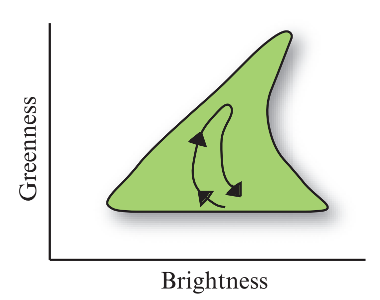
Normalized Difference Moisture or Water Index—NDMI or NDWI

As previously discussed, information about vegetation water content has widespread utility in agriculture, forestry, and hydrology (Galvao et al., 2005). Hardisky et al. (1983) and Gao (1996) found that the *Normalized Difference Moisture* or *Water Index* (NDMI or NDWI) based on Landsat TM near- and middle-infrared bands,

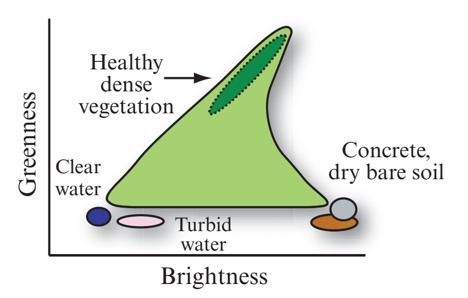
$$\text{NDMI or NDWI} = \frac{(\rho_{nir} - \rho_{swir})}{(\rho_{nir} + \rho_{swir})}, \tag{8.45}$$

was highly correlated with canopy water content and more closely tracked changes in plant biomass and water stress than did the NDVI. Jackson et al. (2004) found that the NDVI saturated while the NDWI continued to document changes in corn and soybean vegetation water content. Jin and Sader (2005) compared the NDMI and the tasseled cap wetness transformation for detecting forest disturbance. Verbesselt et al.

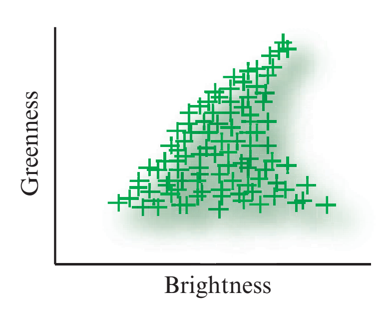
Characteristics of the Kauth-Thomas Tasseled Cap Transformation



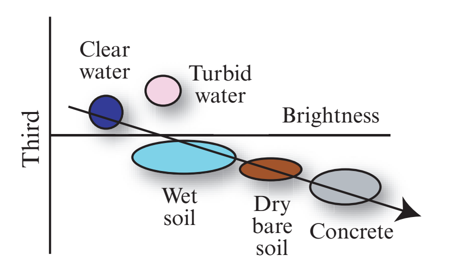
a. Crop development in the tasseled cap brightness-greenness transformation.



b. Location of land cover when plotted in the brightness-greenness spectral space.



c. Actual plot of brightness and greenness values for an agricultural area. The shape of the distribution looks like a cap.



d. Approximate direction of moisture variation in the plane of soils. The arrow points in the direction of less moisture.

FIGURE 8–45 General characteristics of the Kauth-Thomas tasseled cap transformation (based on Crist and Cicone, 1984; Crist and Kauth, 1986). The tasseled cap transformation is applied correctly if the tasseled cap's base is parallel with the brightness axis.

TABLE 8–10 Tasseled cap coefficients for use with Landsat 7 ETM+ data (Huang et al., 2002).

Index	TM1	TM2	TM3	TM4	TM5	TM7
Brightness	.3561	.3972	.3904	.6966	.2286	.1596
Greenness	–.334	–.354	–.456	.6966	–.024	–.263
Wetness	.2626	.2141	.0926	.0656	–.763	–.539
Fourth	.0805	–.050	.1950	–.133	.5752	–.777
Fifth	–.725	–.020	.6683	.0631	–.149	–.027
Sixth	.400	–.817	.3832	.0602	–.109	.0985

(2007) used the NDWI to monitor herbaceous fuel moisture based on SPOT VEGETATION data. Cheng et al. (2008) applied the NDWI to MODIS data and compared the results to AVIRIS-derived canopy equivalent water thickness (EWT).

Perpendicular Vegetation Index—PVI

Richardson and Wiegand (1977) were instrumental in the development of the perpendicular distance to the "soil line" as an indicator of plant development. The "soil line," which is a two-dimensional analog of the Kauth-Thomas soil brightness index, was estimated by linear regression. The *Perpendicular Vegetation Index* (PVI) based on MSS band 4 data was

$$\text{PVI} = \sqrt{(0.355\,MSS4 - 0.149\,MSS2)^2 + (0.355\,MSS2 - 0.852\,MSS4)^2} \quad (8.46)$$

Guyon et al. (2011) used a more generic PVI to monitor leaf phenology in deciduous broadleaf forests where:

$$\text{PVI} = \frac{(NIR - aRed - b)}{\sqrt{1+a^2}} \quad (8.47)$$

and a and b are the slope and the intercept of the soil line, respectively.

Leaf Water Content Index—LWCI

Hunt et al. (1987) developed the *Leaf Water Content Index* (LWCI) to assess water stress in leaves:

$$\text{LWCI} = \frac{-\log[1-(NIR_{TM4} - MidIR_{TM5})]}{-\log[1-(NIR_{TM4_{ft}} - MidIR_{TM5_{ft}})]}, \quad (8.48)$$

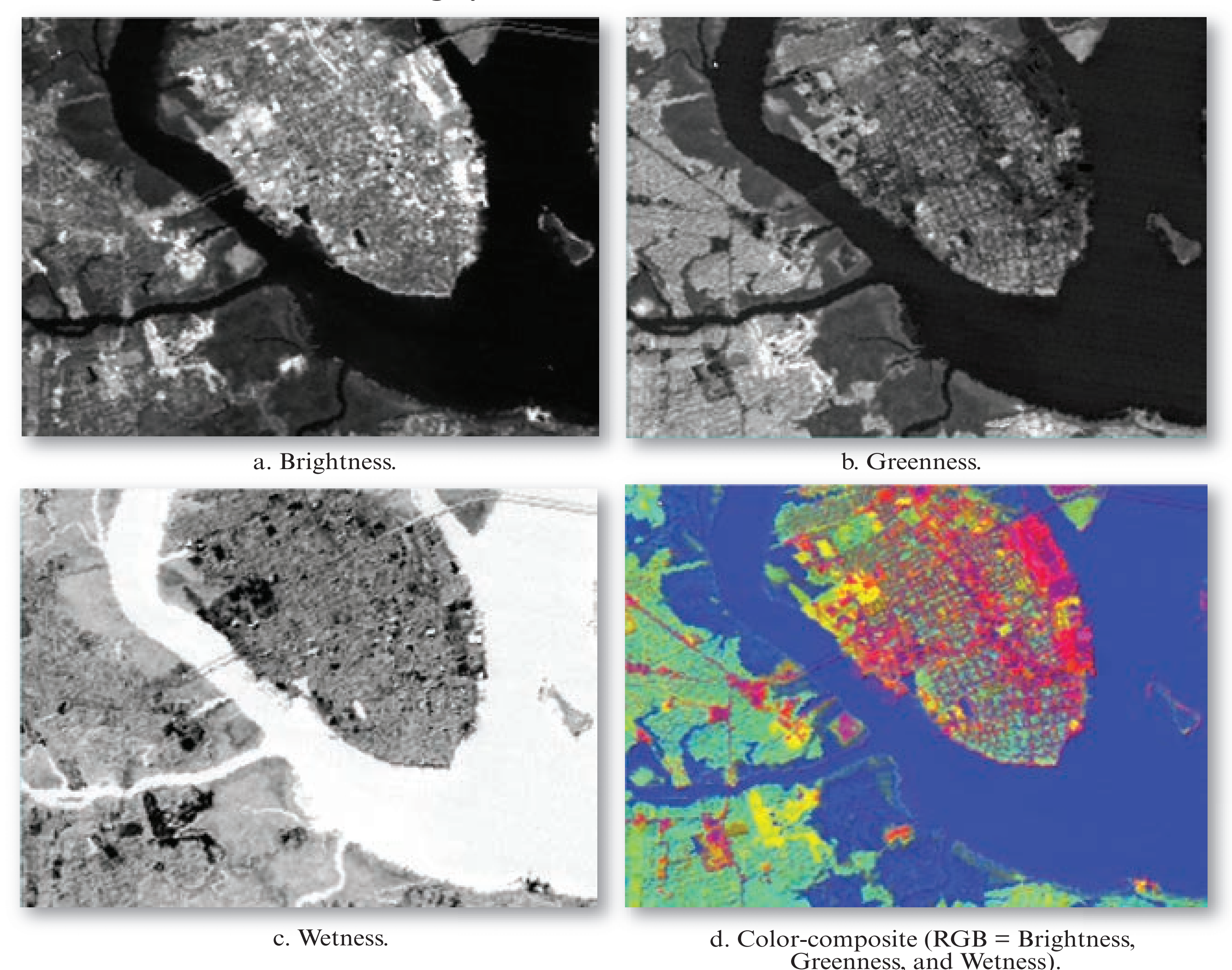


FIGURE 8–46 Brightness, greenness, and wetness images derived by applying Kauth-Thomas tasseled cap transformation coefficients to November 9, 1982, Charleston, SC, Thematic Mapper data (the thermal band was not used). Original imagery courtesy of NASA.

where *ft* represents reflectance in the specified bands when leaves are at their maximum relative water content (fully turgid; RWC) defined as:

$$RWC = \frac{\text{field weight} - \text{oven dry weight}}{\text{turgid weight} - \text{oven dry weight}} \times 100 \quad (8.49)$$

Soil Adjusted Vegetation Index—SAVI

The utility of the NDVI and related indices for satellite and airborne assessment of the Earth's vegetation cover has been demonstrated for almost three decades. The time series analysis of seasonal NDVI data has provided a method of estimating net primary production over varying biome types, of monitoring phenological patterns of the Earth's vegetated surface, and of assessing the length of the growing season and dry-down periods (Huete and Liu, 1994; Ramsey et al., 1995). For example, global vegetation analysis was initially based on linearly regressing NDVI values (derived from AVHRR, Landsat MSS, Landsat TM, and SPOT HRV data) with *in situ* measurements of LAI, APAR, percent cover, and/or biomass. This empirical approach revolutionized global science land-cover biophysical analysis in just one decade (Running et al., 1994). Unfortunately, studies have found that the empirically derived NDVI products can be unstable, varying with soil color and moisture conditions, bidirectional reflectance distribution function (BRDF) effects, atmospheric conditions, and the presence of dead material in the canopy itself (Qi et al., 1995). For example, Goward et al. (1991) found errors of ±50% in NDVI data used for global vegetation studies derived from the NOAA Global Vegetation Index product. What is needed are globally accurate NDVI-related products that do not need to be calibrated by *in situ*

measurement within each geographic area yet will remain constant under changing atmospheric and soil background conditions (Huete and Justice, 1999).

Therefore, emphasis has been given to the development of improved vegetation indices that take advantage of calibrated sensor systems such as the Moderate Resolution Imaging Spectrometer (MODIS) (Running et al., 1994). Although the NDVI has been shown to be useful in estimating vegetation properties, many important external and internal influences restrict its global utility. The improved indices typically incorporate a soil background and/or atmospheric adjustment factor.

The *Soil Adjusted Vegetation Index* (SAVI) is

$$\text{SAVI} = \frac{(\rho_{nir} - \rho_{red})}{(\rho_{nir} + \rho_{red} + L)}(1 + L) \tag{8.50}$$

where L is a canopy background adjustment factor that accounts for differential red and near-infrared extinction through the canopy (Huete, 1988; Huete et al., 1992; Karnieli, et al., 2001). The value of L depends on the proportional vegetation cover as well as the vegetation density. For very sparse vegetation or bare soils, L approximates one, whereas L converges to zero in densely vegetated areas (in this case the SAVI and NDVI become equal) (Huete and Liu, 1994; Sonnenschein et al., 2011). The utility of SAVI for minimizing the soil "noise" inherent in the NDVI has been corroborated in many studies. Qi et al. (1995) developed a modified SAVI, called MSAVI, that uses an iterative, continuous L function to optimize soil adjustment and increase the dynamic range of SAVI.

Atmospherically Resistant Vegetation Index—ARVI

SAVI was made much less sensitive to atmospheric effects by normalizing the radiance in the blue, red, and near-infrared bands. This became the *Atmospherically Resistant Vegetation Index* (ARVI):

$$\text{ARVI} = \left(\frac{\rho_{nir} - \rho_{rb}}{\rho_{nir} + \rho_{rb}}\right) \tag{8.51}$$

where

$$\rho_{rb} = \rho_{red} - \gamma(\rho_{blue} - \rho_{red}).$$

The technique requires prior correction for molecular scattering and ozone absorption of the blue, red, and near-infrared remote sensor data. ARVI uses the difference in the radiance between the blue channel and the red channel to correct the radiance in the red channel and thus reduce atmospheric effects. Unless the aerosol model is known *a priori*, gamma (γ) is normally equal to 1.0 to minimize atmospheric effects. Kaufman and Tanre (1992) provide guidelines about where different gammas might be used over continental, maritime, desert (e.g., the Sahel in Africa), or heavily vegetated areas. Colombo et al. (2003) used ARVI and other VIs with high spatial resolution IKONOS imagery and texture measures to extract leaf-area-index information.

Soil and Atmospherically Resistant Vegetation Index—SARVI

Huete and Liu (1994) integrated the L function from SAVI and the blue-band normalization in ARVI to derive a *Soil and Atmospherically Resistant Vegetation Index* (SARVI) that corrects for both soil and atmospheric noise, as would a modified SARVI (MSARVI):

$$\text{SARVI} = \frac{\rho_{nir} - \rho_{rb}}{\rho_{nir} + \rho_{rb} + L} \tag{8.52}$$

and

$$\text{MSARVI} = \frac{2\rho_{nir} + 1 - \sqrt{\left[(2\rho_{nir} + 1)^2 - \gamma(\rho_{nir} - \rho_{rb})\right]}}{2} \tag{8.53}$$

Huete and Liu (1994) performed a sensitivity analysis on the original NDVI and improved vegetation indices (SAVI, ARVI, SARVI, MSARVI) and drew the following conclusions:

- If there were a total atmospheric correction, then there would mainly be "soil noise," and the SAVI and MSARVI would be the best equations to use, and the NDVI and ARVI would be the worst.
- If there were a partial atmospheric correction to remove the Rayleigh and ozone components, then the best vegetation indices would be the SARVI and MSARVI, with the NDVI and ARVI being the worst.
- If there were no atmospheric correction at all (i.e., no Rayleigh, ozone, or aerosol correction), the SARVI would become slightly worse but still would have the least overall noise. The NDVI and ARVI would have the most noise and error.

Aerosol Free Vegetation Index—AFRI

Karnieli et al. (2001) found that under clear sky conditions the spectral bands centered on 1.6 and 2.1 μm are highly correlated with visible spectral bands centered on blue (0.469 μm), green (0.555 μm), and red (0.645 μm). Empirical linear relationships such as $\rho_{0.469\mu m} = 0.25\rho_{2.1\mu m}$, $\rho_{0.555\mu m} = 0.33\rho_{2.1\mu m}$ and $\rho_{0.645\mu m} = 0.66\rho_{1.6\mu m}$ were found to be statistically significant. Therefore, based on these and other relationships, two

different *Aerosol Free Vegetation Indices* (AFRI) were developed:

$$\mathrm{AFRI}_{1.6\mu m} = \frac{(\rho_{nir} - 0.66\rho_{1.6\mu m})}{(\rho_{nir} + 0.66\rho_{1.6\mu m})} \text{ and} \quad (8.54)$$

$$\mathrm{AFRI}_{2.1\mu m} = \frac{(\rho_{nir} - 0.5\rho_{2.1\mu m})}{(\rho_{nir} + 0.5\rho_{2.1\mu m})}. \quad (8.55)$$

Under clear sky conditions, the AFRIs (especially $\mathrm{AFRI}_{2.1\mu m}$) result in values very similar to NDVI. However, if the atmospheric column contains smoke or sulfates then the AFRIs are superior to NDVIs. This is because the electromagnetic energy centered at 1.6 and 2.1 μm is able to penetrate the atmospheric column better than red wavelength energy used in the NDVI. Therefore, the AFRIs have a major application in assessing vegetation in the presence of smoke, anthropogenic pollution, and volcanic plumes. Limited success is expected in the case of dust due to presence of larger particles that are similar in size to the wavelength, and therefore not transparent at 2.1 μm (Kaufman et al., 2000). The AFRIs can be implemented using any sensor that incorporates bands centered on 1.6 and/or 2.1 μm such as the Landsat TM, and ETM^+, Landsat 8, MODIS, ASTER, the Japanese Earth Resource Satellite-Optical System (JERS-OPS), SPOT 4-Vegetation, and IRS-1C/D.

Enhanced Vegetation Index—EVI

The MODIS Land Discipline Group developed the *Enhanced Vegetation Index* (EVI) for use with MODIS data (Huete et al., 1997; Jiang et al., 2008):

$$\mathrm{EVI} = G\frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + C_1\rho_{red} - C_2\rho_{blue} + \mathrm{L}}. \quad (8.56)$$

where *nir*, *red*, and *blue* are atmospherically corrected or partially atmospherically-corrected (Rayleigh and ozone absorption) surface reflectances. The EVI is a modified NDVI with a soil adjustment factor, *L*, and two coefficients, C_1 and C_2, which describe the use of the blue band in correction of the red band for atmospheric aerosol scattering. The coefficients, C_1, C_2, and *L*, are empirically determined as 6.0, 7.5, and 1.0, respectively. *G* is a gain factor set to 2.5. This algorithm has improved sensitivity to high-biomass regions and has improved vegetation monitoring through a decoupling of the canopy background signal and a reduction in atmospheric influences (Huete et al., 1997; Huete and Justice, 1999; Huete et al., 2002; Wagle et al., 2014). A comparison of MODIS-derived NDVI values and MODIS-derived EVI values for dense vegetation is shown in Figure 8-47 (Didan, 2002). The EVI has improved sensitivity to high biomass. Wardlow et al. (2007) found that as NDVI began to approach an asymptotic level at the peak of the growing season over cropland, EVI exhibited more sensitivity during this growth stage.

Two MODIS vegetation indices, the MODIS NDVI and the MODIS EVI, are produced at 1 × 1 km and 500 × 500 m resolutions (and occasionally at 250 × 250 m) using 16-day compositing periods. A MODIS EVI map of the Earth obtained starting on day 193 of 2003 is shown in Figure 8-48.

EVI2 is a two-band version of EVI. It was developed for use with remote sensing systems that do not have a blue band (Jiang et al., 2008; Liu et al., 2012):

$$\mathrm{EVI2} = 2.5\frac{(\rho_{nir} - \rho_{red})}{(\rho_{nir} + 2.4\rho_{red} + 1)}. \quad (8.57)$$

Jiang et al. (2008) suggest that the differences between EVI and EVI2 are insignificant over a large sample of snow/ice-free land cover types, phenologies, and scales when atmospheric influences are minimal.

Triangular Vegetation Index—TVI

Broge and Leblanc (2000) developed a *Triangular Vegetation Index* (TVI), which describes the radiative energy absorbed by pigments as a function of the relative difference between red and near-infrared reflectance in conjunction with the magnitude of reflectance in the green region, where the light absorption by chlorophyll *a* and *b* is relatively insignificant. The TVI index is calculated as the area of the triangle defined by the green peak, the chlorophyll absorption minimum, and the near-infrared shoulder in spectral space. It is based on the fact that both chlorophyll absorption causing a decrease of red reflectance and leaf tissue abundance causing increased near-infrared reflectance will increase the total area of the triangle. The TVI index encompasses the area spanned by the triangle ABC with the coordinates given in spectral space:

$$\mathrm{TVI} = \frac{1}{2}(120(\rho_{nir} - \rho_{green})) - 200(\rho_{red} - \rho_{green}) \quad (8.58)$$

where ρ_{green}, ρ_{red}, and ρ_{nir} are the reflectances centered at 0.55 μm, 0.67 μm, and 0.75 μm, respectively. Pena-Barragan et al. (2011) applied the index to ASTER optical data for crop identification.

Reduced Simple Ratio—RSR

Chen et al. (2002) modified the simple ratio algorithm to include information from the short-wavelength infrared (SWIR) band found in the SPOT VEGETATION sensor. They used the sensor to map the spatial

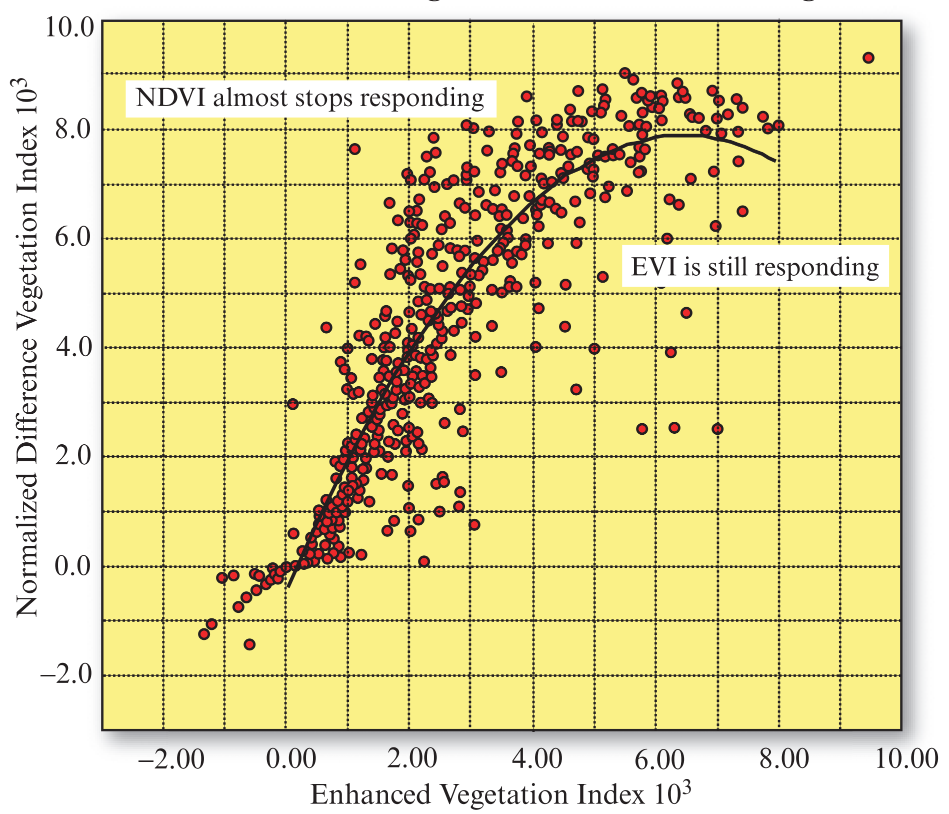


FIGURE 8–47 A comparison of the sensitivity of the MODIS-derived Normalized Difference Vegetation Index (NDVI) values and MODIS Enhanced Vegetation Index (EVI) values for dense vegetation (based on Didan, 2002).

distribution of leaf-area-index in Canada. The *Reduced Simple Ratio* (RSR) is:

$$RSR = \frac{\rho_{nir}}{\rho_{red}}\left(1 - \frac{\rho_{swir} - \rho_{swirmin}}{\rho_{swirmax} + \rho_{swirmin}}\right) \quad (8.59)$$

where $\rho_{swirmin}$ and $\rho_{swirmax}$ are the minimum and maximum SWIR reflectance found in each image and defined as the 1 percent minimum and maximum cut-off points in the histograms of SWIR reflectance in a scene. The major advantages of RSR over SR are a) the difference between land-cover types is much reduced so that the accuracy for leaf-area-index retrieval for mixed cover types can be improved or a single LAI algorithm can be developed without resorting to coregistered land-cover maps as the first approximation, and b) the background (understory, litter, and soil) influence is suppressed using RSR because the SWIR band is most sensitive to the amount of vegetation containing liquid water in the background (Chen et al., 2002). Gray and Song (2012) used the RSR in their mapping of leaf-area-index using spatial, spectral and temporal data from multiple sensors.

Chlorophyll Absorption in Reflectance Index—CARI; Modified Chlorophyll Absorption in Reflectance Index—MTCARI; Optimized Soil-Adjusted Vegetation Index—OSAVI; Ratio TCARI/OSAVI

Many scientists are interested in the amount of chlorophyll in vegetation (e.g., Daughtry et al., 2000). Kim et al. (1994) developed the *Chlorophyll Absorption in Reflectance Index* (CARI). It was modified to become the *Transformed Chlorophyll Absorption in Reflectance Index* (TCARI):

$$TCARI = 3\left[(\rho_{700} - \rho_{670}) - 0.2(\rho_{700} - \rho_{550})\left(\frac{\rho_{700}}{\rho_{670}}\right)\right] \quad (8.60)$$

It uses bands corresponding to the minimum absorption of the photosynthetic pigments, centered at 550 and 700 nm, in conjunction with the chlorophyll *a* maximum absorption band, around 670 nm. The choice of 700 nm is due to its location at the boundary between the region where vegetation reflectance is dominated by pigment absorption and the beginning of the red edge portion where vegetation structural characteristics (i.e., the spongy mesophyll) have more influence on reflectance (Kim et al., 1994).

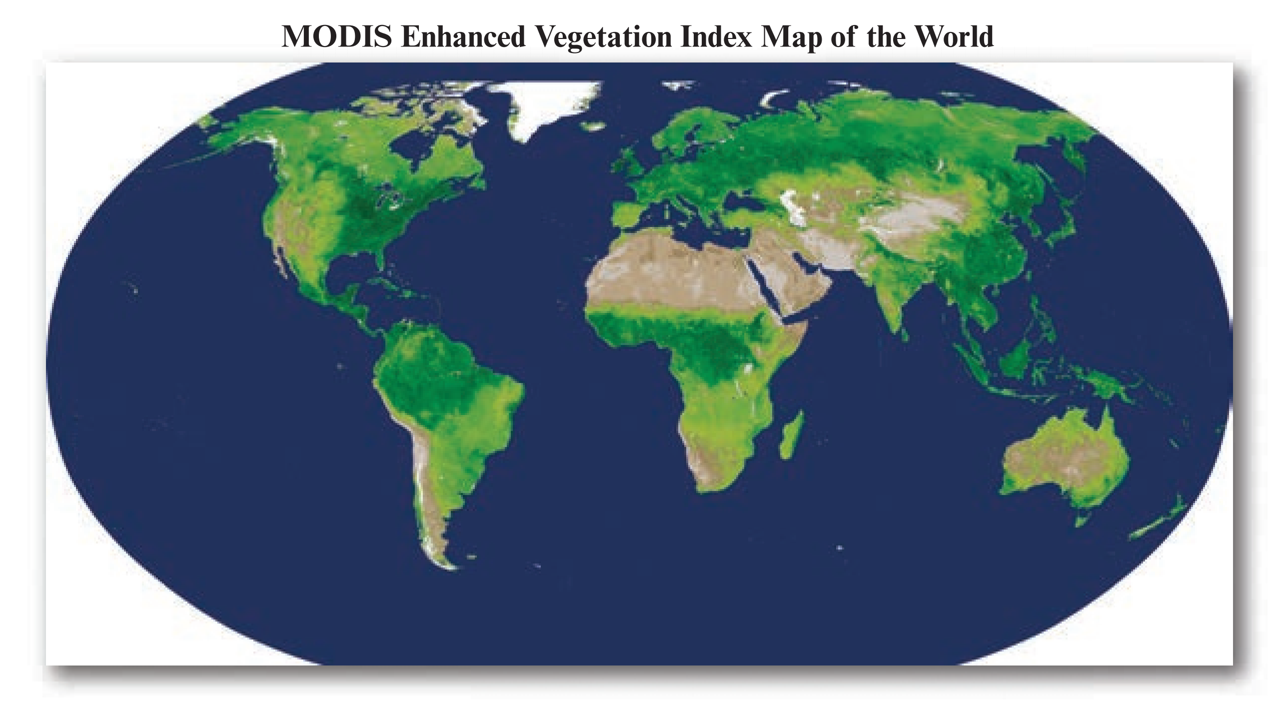


FIGURE 8–48 MODIS Enhanced Vegetation Index (EVI) map of the world obtained over a 16-day period beginning on day 193 of 2003. The greener the area, the greater the amount of biomass (courtesy Terrestrial Biophysics and Remote Sensing MODIS Team, University of Arizona and NASA).

Unfortunately, TCARI is still sensitive to the underlying soil reflectance properties, particularly for vegetation with a low leaf-area-index. Therefore, Daughtry et al. (2000) proposed that TCARI be combined with a generic soil line vegetation index like the *Optimized Soil-Adjusted Vegetation Index* (OSAVI) (Rondeaux et al., 1996):

$$\text{OSAVI} = \frac{(1 + 0.16)(\rho_{800} - \rho_{670})}{(\rho_{800} + \rho_{670} + 0.16)} . \quad (8.61)$$

The ratio became

$$\frac{\text{TCARI}}{\text{OSAVI}} \quad (8.62)$$

and is highly correlated with vegetation chlorophyll content (Haboudane et al., 2002).

Visible Atmospherically Resistant Index—VARI

Many resource managers would like vegetation fraction information (e.g., 60%) (Rundquist, 2002). Building upon the Atmospherically Resistant Vegetation Index, scientists developed the *Visible Atmospherically Resistant Index* (VARI) computed as (Gitelson et al., 2002):

$$\text{VARI}_{green} = \frac{\rho_{green} - \rho_{red}}{\rho_{green} + \rho_{red} - \rho_{blue}} . \quad (8.63)$$

The index is minimally sensitive to atmospheric effects, allowing estimation of vegetation fraction with an error of <10% in a wide range of atmospheric optical thickness.

Normalized Difference Built-Up Index—NDBI

Many professionals working on urban/suburban problems are interested in monitoring the spatial distribution and growth of urban built-up areas. These data can be used for watershed runoff prediction and other planning applications. Zha et al. (2003) calculated a *Normalized Difference Built-up Index* (NDBI):

$$\text{NDBI} = B_u - \text{NDVI} \quad (8.64)$$

where

$$B_u = \frac{NIR_{TM4} - MidIR_{TM5}}{NIR_{TM4} + MidIR_{TM5}} . \quad (8.65)$$

This resulted in an output image that contained only built-up and barren pixels having positive values while all other land cover had a value of 0 or –254. The technique was reported to be 92% accurate.

Vegetation Adjusted Nighttime Light (NTL) Urban Index—VANUI

The Defense Meteorological Satellite Program/Operational Linescan System (DMSP/OLS) Nighttime light (NTL) data are able to provide information on nighttime luminosity which is highly correlated with the existence of urbanized areas and energy consumption. Sometimes the utility of the NTL data is limited due to

saturation of the data values, especially in downtown central business districts. Therefore, Zhang et al. (2013) developed the Vegetation Adjusted NTL Urban Index (VANUI), which incorporates MODIS-derived NDVI information along with NTL data:

$$\text{VANUI} = (1 - \text{NDVI}) \times \text{NTL}\,. \tag{8.66}$$

The VANUI image exhibits reduced NTL saturation and corresponds to both biophysical (e.g., soil, water, vegetation) and urban characteristics.

Red-Edge Position Determination—REP

The abrupt change in the 680 – 800 nm region of reflectance spectra of leaves caused by the combined effects of strong chlorophyll absorption and leaf internal scattering is called the red edge. The *red edge position* (REP) is the point of maximum slope on a vegetation reflectance spectrum between the red and near-IR wavelengths. The red edge was first described by Collins (1978), and is perhaps the most studied feature on the vegetation spectral curve. The REP is useful because it is strongly correlated with foliar chlorophyll content and can be a sensitive indicator of vegetation stress. Determining the red edge position using remote sensing data usually requires the collection of hyperspectral data.

A linear method proposed by Clevers (1994) can be implemented that makes use of four narrow bands and is computed as follows:

$$\text{REP} = 700 + 40\left[\frac{\rho_{(\text{red edge})} - \rho_{(700\text{nm})}}{\rho_{(740\text{nm})} - \rho_{(700\text{nm})}}\right], \tag{8.67}$$

where

$$\rho_{(\text{red edge})} = \frac{\rho_{(670\text{nm})} + \rho_{(780\text{nm})}}{2}\,. \tag{8.68}$$

Dawson and Curran (1998) and Baranoski and Rokne (2005) summarize additional methods used to determine the red edge position.

Photochemical Reflectance Index—PRI

Mand et al. (2010) used a photochemical reflectance index (PRI):

$$\text{PRI} = \frac{(\rho_{531} - \rho_{570})}{(\rho_{531} + \rho_{570})} \tag{8.69}$$

to monitor experimental warming and drought shrub land study areas in Europe. PRI was significantly related to effective green leaf-area-index (LAIe), however, it was found to be very sensitive to soil reflectance properties especially in areas with low green LAIe. Sims et al. (2011) used PRI in their study of view angle effects on MODIS data used in vegetation indices because it has been shown to be highly correlated with vegetation light-use-efficiency at leaf, canopy and stand levels.

NDVI and Cellulose Absorption Index (CAI) to Quantify Fractional Cover of Photosynthetic Vegetation, Non-Photosynthetic Vegetation and Bare Soil

Guerschman et al. (2009) state that quantitative estimation of fractional cover of photosynthetic vegetation (f_{PV}), non-photosynthetic vegetation which includes dead plant matter (f_{NPV}), and bare soil (f_{BS}) is critical for natural resource management and for modeling carbon dynamics. They developed a methodology based on the use of NDVI and a Cellulose Absorption Index (CAI) that can be used together in a linear unmixing approach to identify the characteristics of the fractional cover. The CAI is computed using special short wavelength infrared (*swir*) regions of the electromagnetic spectrum:

$$\text{CAI} = [0.5 \times (\rho_{2.0} + \rho_{2.2}) - \rho_{2.1}] \times 10 \tag{8.70}$$

where $\rho_{2.0}$ = EO-1 Hyperion hyperspectral bands in the region from 2,022 – 2,032 nm; $\rho_{2.1}$ = bands in the region from 2,093 – 2,113 and ρ_{22} = bands in the region from 2,184 – 2,204 (Guerschman et al., 2009). Fractional is estimated by determining where a pixel resides in the NDVI_CAI spectral feature space shown in Figure 8-49. The three endpoints in the feature space represent pure “endmembers” associated with bare soil (f_{BS}), non-photosynthetic vegetation (f_{NPV}), and photosynthetic vegetation (f_{PV}). Reflectance spectra that are mixtures (e.g., at locations 1 to 4) of the three pure endmembers will lie within the yellow NDVI_CAI feature space. In this hypothetical example, a pixel with #1 characteristics will likely consist of 50% bare soil and 50% photosynthetic vegetation. A pixel with #4 characteristic will likely have equal proportions (e.g., 33%) of each of the three endmembers.

MERIS Terrestrial Chlorophyll Index (MTCI)

Dash et al. (2010) used the MERIS Terrestrial Chlorophyll Index (MTCI) to map the spatio-temporal variation of vegetation phenology in India using the following equation:

$$\text{MTCI} = \frac{\rho_{band\ 10} - \rho_{band\ 9}}{\rho_{band\ 9} - \rho_{band\ 8}} = \frac{\rho_{753.75} - \rho_{708.75}}{\rho_{708.75} - \rho_{681.25}} \tag{8.71}$$

and is based on MERIS bands 8, 9, and 10. Their study used 8-day temporal composites of MERIS derived MTCI data. They suggest that the MTCI combines information on LAI and the chlorophyll concentration of leaves to produce an image of chlorophyll content. The MTCI is an operational European Space Agency Level 2 land product.

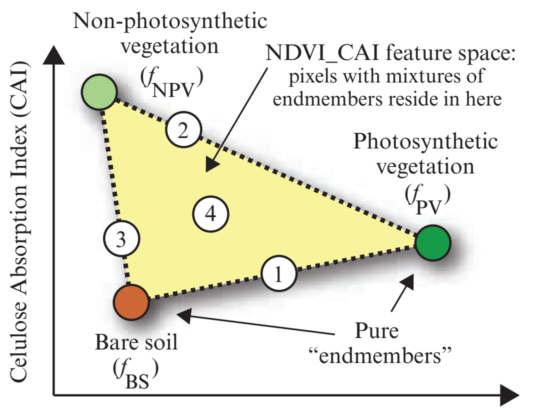


FIGURE 8–49 Gureschman et al. (2009) suggest that it is possible to quantify fractional cover by analyzing information in the spectral feature space bounded by information from the Cellulose Absorption Index (CAI) and the Normalized Difference Vegetation Index (NDVI). Reflectance spectra that are mixtures (e.g., at locations 1-4) of the three pure endmembers will lie within the yellow feature space (based on Guerschman et al., 2009).

Normalized Burn Ratio (NBR)

One of the most widely used spectral indices for mapping burn severity is the Normalized Burn Ratio (NBR) (Brewer et al., 2005). The NBR combines the reflectances in the near-infrared (ρ_{nir}) and the short-wavelength infrared bands (ρ_{swir}) (Boer et al., 2008):

$$NBR = \frac{(\rho_{nir} - \rho_{swir})}{(\rho_{nir} + \rho_{swir})} \qquad (8.72)$$

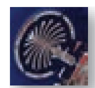
The NBR is sometimes used to map fire-affected areas using a single post-fire image, but is more commonly applied to quantify burn severity as the difference (Δ) between pre- and post-fire values observed in registered multiple-date images (Boer et al., 2008):

$$\Delta NBR = NBR_{pre-fire} - NBR_{post-fire} \qquad (8.73)$$

The NBR is used in the USDA Forest Service FIREMON system for the assessment and monitoring of fire impacts (Lutes et al., 2006; Boer et al., 2008).

Vegetation Suppression

Vegetation suppression algorithms model the amount of vegetation per pixel using a vegetation transform such as NDVI. The model calculates the relationship of each input band with vegetation, then it decorrelates the vegetative component of the total signal on a pixel-by-pixel basis for each band (Crippen and Blom, 2001; ENVI Suppression, 2013). Vegetation suppression is most commonly used in lithologic mapping and linear feature enhancement in areas with open canopies. This method helps analysts better interpret geologic and urban features and works best with open-canopy vegetation in medium spatial resolution (e.g., 30 × 30 m) imagery. For closed canopies in moderate-resolution data, vegetation suppression is primarily used for linear feature enhancement.



Texture Transformations

When humans visually interpret remotely sensed imagery, they synergistically take into account context, edges, texture, and tonal variation or color. Conversely, most digital image processing classification algorithms are based only on the use of the spectral (tonal) information (i.e., brightness values). Thus, it is not surprising that there has been considerable activity in trying to incorporate some of these other characteristics into digital image classification procedures.

A *discrete tonal feature* is a connected set of pixels that all have the same or almost the same gray shade (brightness value). When a small area of the image (e.g., a 3 × 3 area) has little variation of discrete tonal features, the dominant property of that area is a gray shade. Conversely, when a small area has a wide variation of discrete tonal features, the dominant property of that area is *texture.* Most researchers trying to incorporate texture into the classification process have attempted to create a new texture image that can then be used as another feature or band in the classification process. Thus, each new pixel of the texture image has a brightness value that represents the texture at that location (i.e., $BV_{i,j,\text{texture}}$).

First-Order Statistics in the Spatial Domain

One class of picture properties that can be used for texture synthesis is first-order statistics of local areas such as mean Euclidean distance, variance, skewness, kurtosis (e.g., Luo and Mountrakis, 2010; Warner, 2011; Culbert et al., 2012; ERDAS, 2013). Typical algorithms include:

$$\text{Mean Euclidean Distance} = \frac{\sum\left[\Sigma_{\lambda}(BV_{c\lambda} - BV_{ij\lambda})^2\right]^{\frac{1}{2}}}{n-1} \qquad (8.74)$$

where

$BV_{c\lambda}$ = brightness value of a window's center pixel for spectral band λ

$BV_{ij\lambda}$ = brightness value for pixel (*i,j*) in spectral band λ of a multispectral image

n = number of pixels in the window.

$$\text{Variance} = \frac{\sum (BV_{ij} - \mu)^2}{n-1} \tag{8.75}$$

where

BV_{ij} = brightness value of pixel (*i,j*)

n = number of pixels in the window

μ = mean of the moving window where

$$\mu = \left(\sum BV_{ij}\right) / n .$$

$$\text{Skewness} = \frac{\left|\sum (BV_{ij} - \mu)^3\right|}{(n-1)(V)^{\frac{3}{2}}} \tag{8.76}$$

where V = Variance.

$$\text{Kurtosis} = \frac{\sum (BV_{ij} - \mu)^4}{(n-1)(V)^2} \tag{8.77}$$

The pixel windows typically range from 3 × 3 to 5 × 5 to 7 × 7. The application of a 3 × 3 Variance texture measure to the near-infrared band of a high spatial resolution digital frame camera image is shown in Figure 8-50b. The brighter the pixel, the greater the heterogeneity (more coarse the texture) within the window. A composite of Variance derived from the NIR, red, and green bands is shown in Figure 8-50c. A color composite of Skewness within the 3 × 3 spatial moving window derived from the NIR, red, and green bands is shown in Figure 8-50d.

Numerous scientists have evaluated these and other texture transformations. For example, Mumby and Edwards (2002) extracted the variance from IKONOS imagery using a 5 × 5 pixel window. They found that when texture features were used in conjunction with spectral data the thematic map accuracy increased for medium and fine levels of coral reef habitat. Ferro and Warner (2002) used a variance texture measure and found that land-cover class separability increased when texture was used in addition to spectral information, and that texture separability increased with larger windows. Wang et al. (2004) used variance texture along with other second-order texture measures in their study of mangroves. Estes et al. (2008) used mean and standard deviation texture and Estes et al. (2010) used standard deviation texture in their analysis of forest antelope habitat using SPOT and ASTER data. Luo and Mountrakais (2010) used the mean and variance in their impervious surface study. Warner (2011) reviewed the characteristics of numerous first-order texture measures. Dronova et al. (2012) used standard deviation texture in their object-based image analysis of wetland plant function types. Gray and Song (2012) used simple variance texture in their studies of LAI extraction from MODIS data.

Edge-Preserving Spectral-Smoothing (EPSS) Variance Texture

Laba et al. (2010) created an iterative, edge-preserving spectral-smoothing (EPSS) texture measure associated with the mean and variance in a 5 × 5 spatial moving window (Figure 8-51) that was applied to individual bands of multispectral IKONOS imagery. The goal was to determine the mean of the central pixel, e.g., at location 23 in Figure 8-51a. The was performed by first computing the variance for a block of nine pixels centered about a given pixel and for blocks of seven pixels that extend from the central pixel in eight directions about the central pixel as shown in Figure 8-51. The block with the *minimum* variance was then identified. Therefore, for each sub-window, k, the mean (Laba et al., 2010):

$$\mu_k = \frac{1}{N}\sum_{i=1}^{3}\sum_{j=1}^{3} x_{i,j} \times m_{i,j,k} \tag{8.78}$$

and variance were computed

$$\sigma_k = \frac{1}{N}\sum_{i=1}^{3}\sum_{j=1}^{3} (x_{i,j} - \mu_k) \times m_{i,j,k} \tag{8.79}$$

where $x_{i,j}$ was the brightness value at the *i,j* location within the kth sub-window, and $m_{i,j,k}$ was a binary mask value (0 or 1) that corresponded with the kth sub-window. The *mean* of the minimum variance sub-window was then assigned to the center pixel, x_c, where

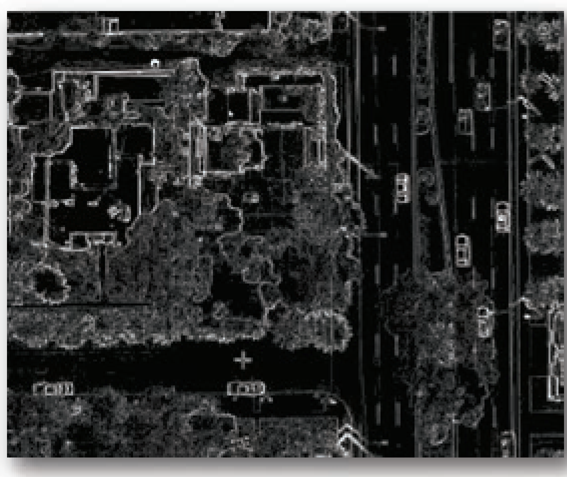
$$x_c = \mu_{min} \text{ such that } \sigma_{min} = \min(\sigma_k). \tag{8.80}$$

The method was applied to all the pixels in each band iteratively until a specified threshold was reached. Laba et al. (2010) extended the method to include information on all the multispectral bands in the dataset. They found that this texture measure preserved abrupt

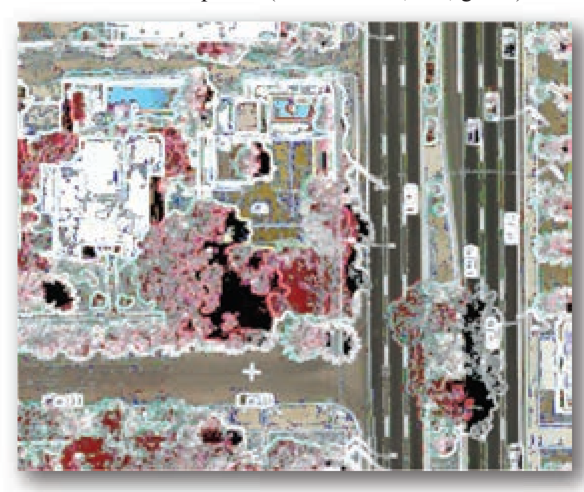
First-order Texture Information Extracted from Digital Frame Camera Imagery



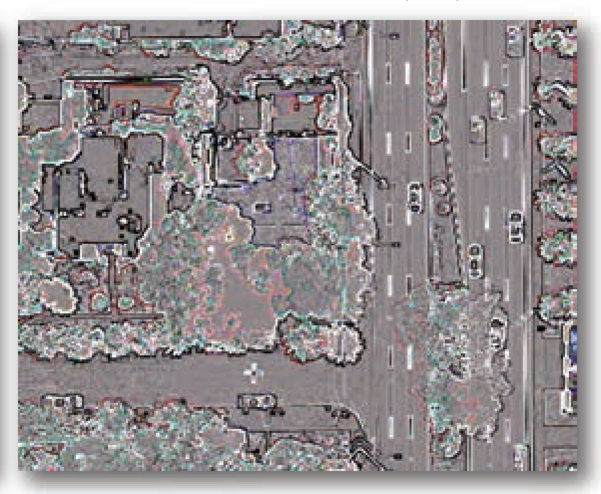
a. Color composite (RGB = NIR, red, green).



b. Variance of the near-infrared (NIR) band.



c. Color composite (RGB = NIR variance, red variance, green variance).



d. Color composite (RGB = NIR skewness, red skewness, green skewness).

FIGURE 8–50 Selected first-order texture measures derived using a 3 × 3 spatial moving window. a) Original image (original aerial photograph courtesy of John Copple, Sanborn Map Company). b) Variance of the near-infrared band. c) Composite of Variance derived from NIR, red, and green bands. d) Composite of Skewness derived from NIR, red, and green bands.

spatial changes in spectral intensity or spectral shape found in the original data.

Conditional Variance Detection

Zhang (2001) developed a texture measure to identify tree features in high spatial resolution imagery. It consists of two components: 1) directional variance detection, and 2) standard local variance measurement, as previously discussed. The spatial logic of the directional variance detection is shown in Figure 8-52. Directional variance detection is used to identify whether the central pixel within the moving spatial window (e.g., 7 × 7) is located in a "treed" area. If the central pixel is in a treed area, the local variance calculation is then carried out to highlight the pixel in an output file. Otherwise, the local variance calculation is avoided to suppress the pixel. To effectively detect edges of other objects and separate them from treed areas, the size of the window for directional variance detection should be larger (e.g., 7 × 7) than the window for local variance calculation (e.g., 3 × 3).

The directional variance detector measures the pixel variances along the shaded pixels in Figure 8-52 on each side of the central pixel in four directions using the equation:

$$D_{var} = \frac{1}{n}\sum_{i=-n}^{n-1}[f(i,j) - \overline{f(i,j)}]^2 \tag{8.81}$$

with $\overline{f(i,j)} = \frac{1}{n}\sum_{i=-n}^{n-1} f(i,j)$

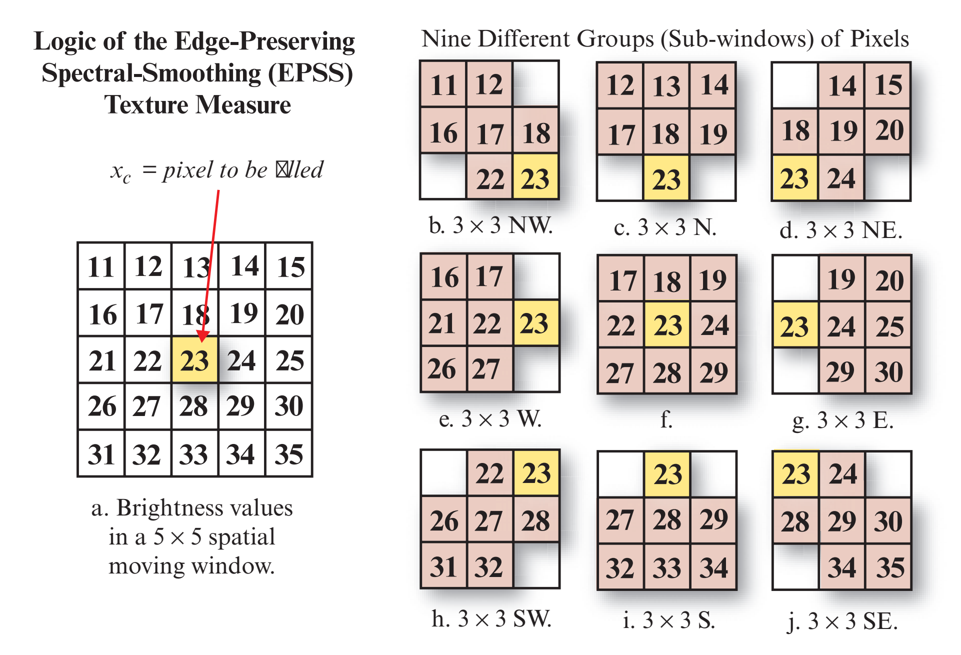


FIGURE 8–51 The logic of the Edge-Preserving Spectral-Smoothing (EPSS) texture measure. a) The 5 × 5 pixel spatial moving window with the pixel to be filled identified. b-j) Nine different groups of pixels are evaluated to come up with the *mean* of the minimum-variance sub-window which is then assigned to x_c. Eight of the groups are associated with the cardinal directions. The method can be extended to work with multispectral data (based on Laba et al., 2010).

and

$i < 0, j = 0$ for the upper side

$i \geq 0, j = 0$ for the lower side

$i = 0, j < 0$ for the left side

$i \geq 0, j \geq 0$ for the right side

where D_{var} is directional variance, $f(i,j)$ is the value of the pixel located at the ith row and jth column in the spatial moving window (Figure 8-52), and n is the pixel count on each side of the central pixel.

If the variance on one of the four sides is less than a user-specified threshold, it is concluded that there is a homogenous area on this side or the central line is along a straight edge. In this case, the central pixel is regarded as a non-tree pixel. It is assigned a lower value and the local variance calculation is not carried out. If, on the other hand, the variance in one of the directions exceeds the threshold, then the calculation of the local variance (e.g., within a 3 × 3 window) is carried out and the central pixel receives a higher value. Zhang (2001) determined that the algorithm detected tree features more accurately than gray-level co-occurrence matrix-derived texture algorithms, to be discussed.

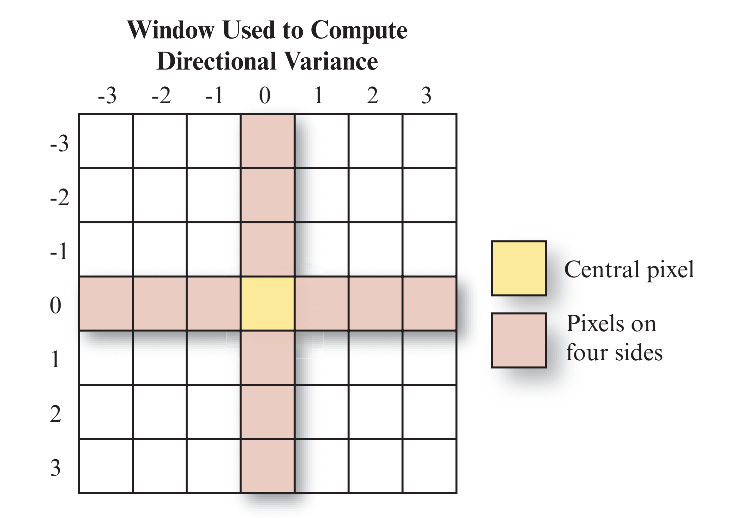
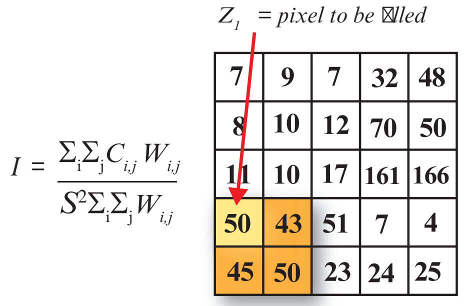


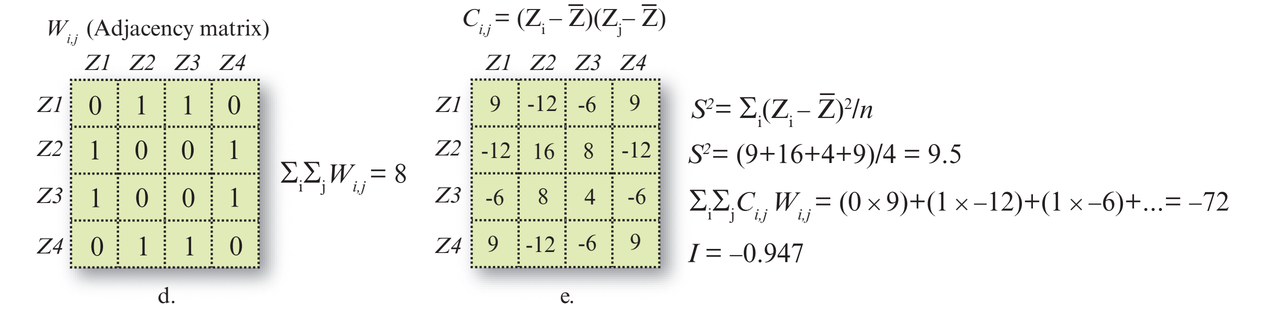
FIGURE 8–52 Window used to compute directional variance (based on Zhang, 2001). It does not have to be a 7 × 7 window.

Min–Max Texture Operator

Briggs and Nellis (1991) developed a *min–max texture operator* based on the analysis of the brightness values found within the following five-element spatial moving window:

Moran's *I* Calculation





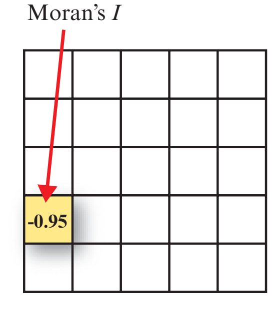
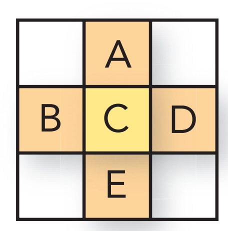


FIGURE 8–53 Hypothetical example of computing Moran's *I* to derive texture in a single band of imagery (based on Purkis et al., 2006).



where

$$\text{Texture}_C = \text{brightest}_{A,\,B,\,C,\,D,\,E} - \text{darkest}_{A,\,B,\,C,\,D,\,E} \quad (8.82)$$

They found that the min–max texture features and NDVI transformations of seven SPOT HRV scenes provided accurate information on the seasonal variation and heterogeneity of a portion of the tallgrass Konza Prairie Research Natural Area in Kansas.

Moran's *I* Spatial Autocorrelation as a Texture Measure

Purkis et al. (2006) used the Moran's *I* spatial autocorrelation metric as a texture operator in their coral reef mapping research. Moran's *I* was calculated using (Purkis et al., 2006):

$$\text{Morans I} = \frac{\sum_i \sum_j C_{i,j} W_{i,j}}{S^2 \sum_i \sum_j W_{i,j}} \quad (8.83)$$

where $W_{i,j}$ is the weight at distance d so that $W_{i,j} = 1$ if point j is within distance d from point i; otherwise, $W_{i,j} = 0$; C_{ij} are deviations from the mean (i.e., $C_{i,j} = (Z_i - \bar{Z})(Z_j - \bar{Z})$ where $\bar{Z}$ is the mean brightness value in the local window. S^2 is the mean of the major diagonal of C_{ij}. Moran's I is computed within a user-specified moving window (e.g., 2 × 2, 3 × 3). An

Hypothetical Example of Creating a Gray-level Co-occurence Matrix (GLCM)

a. Original image =

0	1	1	2	3
0	0	2	3	3
0	1	2	2	3
1	2	3	2	2
2	2	3	3	2

b. $h_{left\ of}$ =

	0	1	2	3
0	1	2	1	0
1	0	1	3	0
2	0	0	3	5
3	0	0	2	2

c. $h_{right\ of}$ =

	0	1	2	3
0	1	0	0	0
1	2	1	0	0
2	1	3	3	2
3	0	0	5	2

d. $h_{symmetric}$ =

	0	1	2	3
0	2	2	1	0
1	2	2	3	0
2	1	3	6	7
3	0	0	7	4

e. $h_{normalized}$ =

	0	1	2	3
0	2/40	2/40	1/40	0/40
1	2/40	2/40	3/40	0/40
2	1/40	3/40	6/40	7/40
3	0/40	0/40	7/40	4/40

f. h_c =

	0	1	2	3
0	.05	.05	.025	0
1	.05	.05	.075	0
2	.025	.075	.15	.175
3	0	0	.175	.01

FIGURE 8–54 Hypothetical example of computing the elements using a 5 × 5 kernel applied to single band of imagery. a) Original brightness values in the 5 × 5 area. b) Number of times brightness value *i* occurs to the left of brightness value *j*. c) Number of times brightness value *i* occurs to the right of brightness value *j*. d) Symmetric. e) Normalization of the matrix. f) GLCM matrix ready for use with algorithms.

example of how Moran's *I* is calculated using hypothetical data in a simple 2 × 2 spatial moving window is shown in Figure 8-53 (Purkis et al., 2006). Moran's *I* varies from +1.0 for positive correlation (a clumped pattern) to –1.0 for negative correlation. If positive autocorrelation exists, pixels with similar spectral characteristics can be inferred to be present within the window, i.e., homogeneous texture is present. If spatial autocorrelation is weak or nonexistent, adjacent pixels within the window have dissimilar spectral values, i.e., heterogenous texture is present (Purkis et al., 2006).

Second-Order Statistics in the Spatial Domain

A suite of very useful texture measures was originally developed by Haralick and associates (Haralick et al., 1973; Haralick and Shanmugan, 1974; Haralick, 1979; Haralick and Fu, 1983; Haralick, 1986). The higher-order set of texture measures is based on brightness value spatial-dependency gray-level co-occurrence matrices (GLCM). The GLCM-derived texture transformations continue to be very widely adopted by the remote sensing community and are often used as an additional feature in multispectral classification (e.g.; Maillard, 2003; Hall-Beyer, 2007; Schowengerdt, 2007; Jensen et al., 2009; Luo and Mountrakis, 2010; Pena-Barragan et al., 2011; Warner, 2011; Culbert et al., 2012; Wang and Zhang, 2014).

But how are these higher-order texture measures computed? If $c = (\Delta x, \Delta y)$ is considered to be a vector in the (x, y) image plane, for any such vector and for any image $f(x, y)$ it is possible to compute the joint probability density of the pairs of brightness values that occur at pairs of points separated by c. If the brightness values in the image can take upon themselves any value from 0 to the highest quantization level in the image (e.g., $quant_k = 255$), this joint density takes the form of an array h_c, where $h_c(i, j)$ is the probability that the pairs of brightness values (i, j) occur at separation c. This array h_c is $quant_k$ by $quant_k$ in size. It is easy to compute the h_c array for $f(x, y)$, where Δx and Δy are integers by simply counting the number of times each pair of brightness values occurs at separation $c = (\Delta x$ and $\Delta y)$ in the image. For example, consider the simple image that has just five lines and five columns and contains brightness values ranging from only 0 to 3 in Figure 8-54a. If (Δx and Δy) = (1, 0), then the numbers are represented by the brightness value spatial-dependency matrix $h_{left\ of}$ shown in Figure 8-54b where the entry in row *i* and column *j* of the matrix is the number of times brightness value *i* occurs to the left of brightness value *j*. For example, brightness value 1 is to the left of brightness value 2 a total of three times (i.e., $h_c[1, 2] = 3$ in the GLCM).

However, in order to produce a "symmetric" GLCM, it is good practice to create a GLCM that also tracks the number of times value *i* occurs to the *right* of brightness value *j* as shown in Figure 8-54c. These two GLCMs are then added together to create the symmetric GLCM (Hall-Beyer, 2007) shown in Figure 8-54d. The GLCM is then normalized by dividing every cell

in it by the total number of possible values. In this case every cell in the matrix is divided by 40. The normalized, symmetrical GLCM (h_c) is now ready to be used in the various texture measures.

It is assumed that all textural information is contained in the brightness value spatial-dependency matrices that are developed for angles of 0°, 45°, 90°, and 135° (Figure 8-55). Generally, the greater the number of values found in the diagonal of the co-occurrence matrices, the more homogeneous the texture is for that part of the image being analyzed.

There are a variety of measures that can be used to extract useful textural information from the h_c matrices. Some weight the off-diagonal elements in the h_c matrices more heavily than the diagonal elements (e.g., Hall-Beyer, 2007). Several of the more widely used GLCM texture measures are the angular second moment (ASM), contrast (CON), correlation (COR), dissimilarity (DIS), entropy (ENT_2) and homogeneity (HOM) (Haralick, 1986; Gong et al., 1992; Zhang, 2001; Maillard, 2003; Kayitakire et al., 2006; Luo and Mountrakis, 2010; Warner, 2011; Wood et al., 2012):

$$\text{ASM} = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} h_c(i,j)^2 \qquad (8.84)$$

$$\text{CON} = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} (i-j)^2 \times h_c(i,j)^2 \qquad (8.85)$$

$$\text{COR} = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} \frac{(i-\mu)(j-\mu)h_c(i,j)^2}{\sigma^2} \qquad (8.86)$$

$$\text{DIS} = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} h_c(i,j)|i-j| \qquad (8.87)$$

$$\text{ENT}_2 = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} h_c(i,j) \times \log[h_c(i,j)] \qquad (8.88)$$

$$\text{HOM} = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} \frac{1}{1+(i-j)^2} \cdot h_c(i,j) \qquad (8.89)$$

where

$quant_k$ = quantization level of band k (e.g., 2^8 = 0 to 255)
$h_c(i, j)$ = the (i, j)th entry in one of the angular brightness value spatial-dependency matrices,

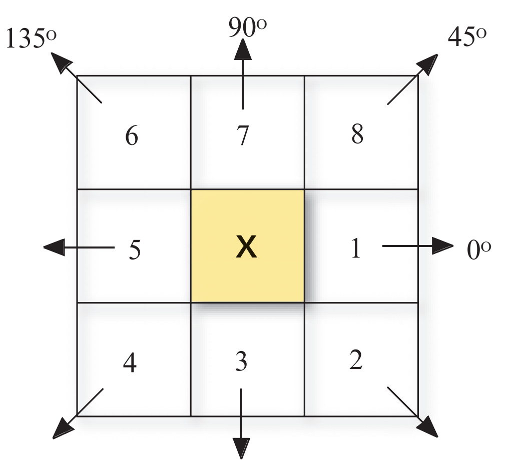


FIGURE 8–55 The eight nearest neighbors of pixel X according to angle ϕ used in the creation of spatial dependency matrices for the measurement of image texture.

and

$$\text{MEAN }(\mu) = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} i \times h_c(i,j) \qquad (8.90)$$

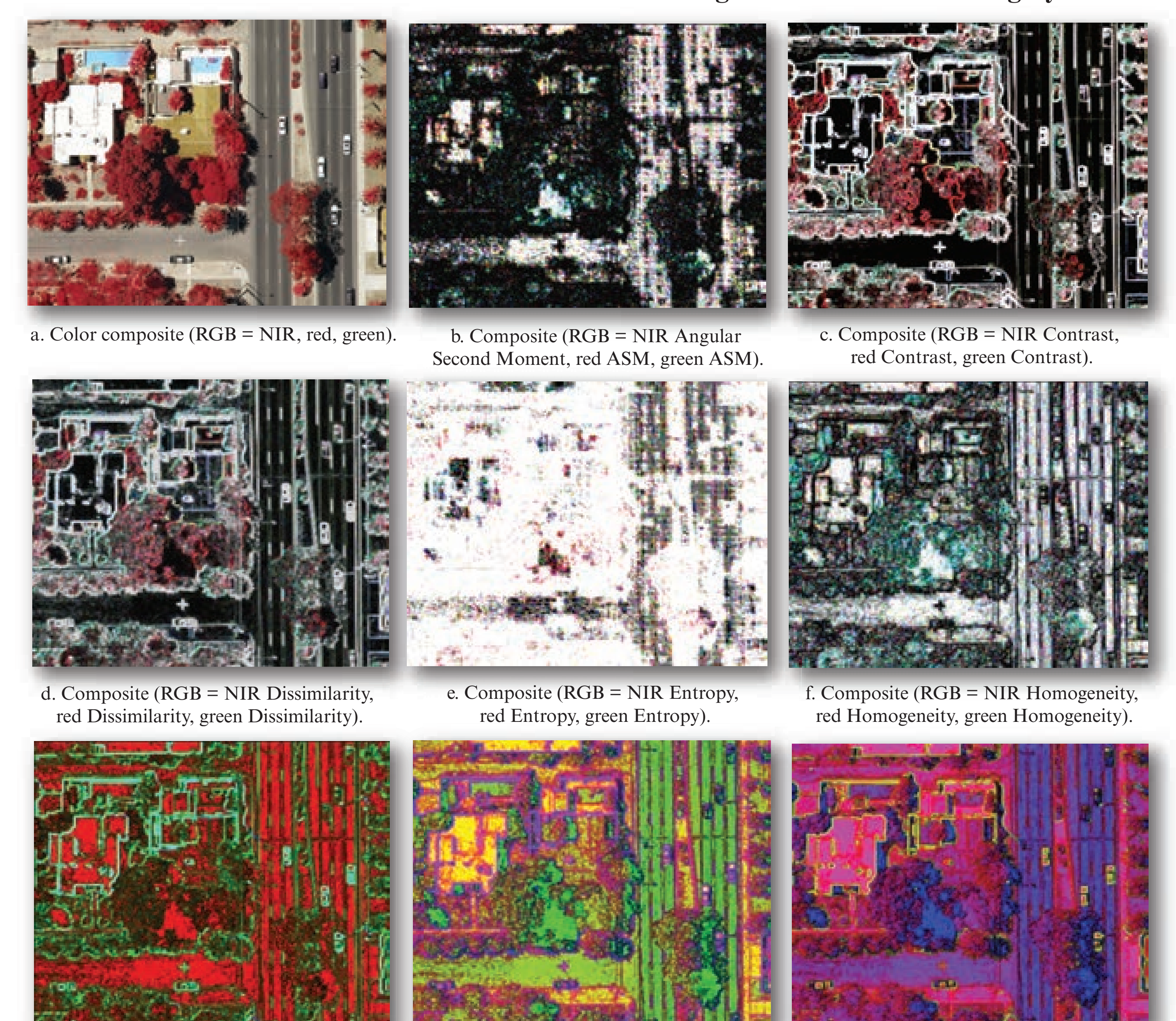
$$\text{VARIANCE }(\sigma^2) = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} (i-\mu)^2 \times h_c(i,j)\,. \qquad (8.91)$$

Note that the MEAN (μ) and VARIANCE (σ^2), which are also texture measures, use the elements of the GLCM rather than the original brightness values in the image during computation (Hall-Beyer, 2007). Four brightness value spatial-dependency matrices (0°, 45°, 90°, and 135°) can be derived for each pixel based on neighboring pixel values. The average of these four measures can be output as the texture value for the pixel under consideration. Most analysts, however, use just one direction (e.g., horizontal or vertical) to create the GLCMs used with the various algorithms.

To create a GLCM-derived texture measure, the analyst usually has to make several important decisions, including (Warner, 2011):

- the texture measure(s),
- window size (e.g., 3 × 3, 5 × 5, 32 × 32)
- input channel (i.e., the spectral channel used to extract texture information)
- quantization level of the input data (e.g., 8-bit, 6-bit, 4-bit) used to produce the output texture image, and
- the spatial component (i.e., the interpixel distance and angle used during co-occurrence computation).

It is possible to use any GLCM-derived texture measure alone. However, Clausi (2002) found that there



a. Color composite (RGB = NIR, red, green).

b. Composite (RGB = NIR Angular Second Moment, red ASM, green ASM).

c. Composite (RGB = NIR Contrast, red Contrast, green Contrast).

d. Composite (RGB = NIR Dissimilarity, red Dissimilarity, green Dissimilarity).

e. Composite (RGB = NIR Entropy, red Entropy, green Entropy).

f. Composite (RGB = NIR Homogeneity, red Homogeneity, green Homogeneity).

g. Composite (RGB = NIR Homogeneity, red Dissimilarity, green Contrast).

h. Composite (RGB = NIR Mean, red Homogeneity, green Dissimilarity).

i. Composite (RGB = NIR Mean, red Homogeneity, green Dissimilarity).

FIGURE 8–56 Selected texture measures derived from gray-level co-occurrence matrices (GLCM). a) Original image. b) Angular Second Moment. c) Contrast. d) Dissimilarity. e) Entropy. f) Homogeneity. g) Composite of Homogeneity, Dissimilarity, and Contrast. h) Composite of Mean, Homogeneity, and Dissimilarity. i) Composite of Mean, Contrast and Dissimilarity (original aerial photography courtesy of John Copple, Sanborn Map Company).

was a preferred subset of statistics (contrast, correlation, and entropy) that was superior to the use of any single statistic or using the entire set of statistics. Gong et al. (1992) found that 3 × 3 and 5 × 5 windows were generally superior to larger windows. Conversely, Ferro and Warner (2002) found that land-cover texture separability increased with larger windows. Several scientists suggest that it is a good idea to reduce the quantization level of the input data (e.g., from 8-bit data with values from 0 to 255 to 5-bit data with values from 0 to 31) when creating the texture images so that the spatial-dependency matrices to be computed for each pixel do not become too large (Clausi, 2002).

Examples of several GLCM-based texture measures are shown in Figure 8-56b-f. The texture images shown were derived for *each* of the three bands in the dataset (green, red, and NIR). It is clear from these examples that the individual texture measures yield different texture information. Additional insight is gained when

color-composites are based on the use of separate texture measures from the three bands, e.g., homogeneity, dissimilarity, and contrast in Figure 8-56g, and mean, homogeneity, and dissimilarity in Figure 8-56h, and mean, contrast, and homogeneity in Figure 8-56i.

Herold et al. (2003) investigated a variety of landscape ecology spatial metrics and GLCM texture measures to classify urban land use in IKONOS high spatial resolution imagery. Wang et al. (2004) used the GLCM texture measures recommended by Clausi (2002) in their study of mangroves. Tuominen and Peekarinen (2005) used five GLCM texture measures applied to aerial photography to inventory forests. Frohn and Hao (2006) used GLCM-derived texture in their evaluation of landscape metric performance using Landsat data of Rondonia, Brazil. Kayitakire et al. (2006) used GLCM Variance, CON, and COR texture measures with four directions and three window sizes (5 × 5, 15 × 15, and 25 × 25) in their study of forest texture using high spatial resolution IKONOS-2 imagery. Hall-Beyer (2007) provides a very useful tutorial on how to compute GCLM-based texture measures and discusses some of the important decisions that must be made when using them. Estes et al. (2008) used the ASM in their analysis of forest antelope habitat using SPOT and ASTER data. Culbert et al. (2012) had good success using GLCM texture measures applied to 114 Landsat scenes when modeling broad-scale patterns of avian species richness across the Midwestern U.S. Luo and Mountrakais (2010) used CON, Energy, and HOM in their impervious surface study. Warner (2011) summarized the characteristics of many GLCM texture measures and suggested ways to overcome challenges of selecting the texture metric, kernel size, spectral band, radiometric quantization, displacement, and angle. Pena-Barragan et al. (2011) used HOM, and ENT in their object-based crop identification. Dronovia et al. (2012) used ENT and HOM texture metrics in their object-based image analysis of wetland plant function types. Wood et al. (2012) used eight GLCM texture measures in their study of grassland-savannah-woodland vegetation structure using aerial photographs and Landsat TM data. Zhang and Xie (2012) investigated the use of GLCM-based texture in their object-based analysis of hyperspectral data of the Florida Everglades.

Texture Units as Elements of a Texture Spectrum

Wang and He (1990) computed texture based on an analysis of the eight possible clockwise ways of ordering the 3 × 3 matrix of pixel values shown in Figure 8-57a. This represents a set containing nine elements $V = \{V_0, V_1, \ldots, V_8\}$, with V_0 representing the brightness value of the central pixel and V_i the intensity of the neighboring pixel i. The corresponding *texture unit* is a set containing eight elements, $\mathrm{TU} = \{E_1, E_2, \ldots, E_8\}$, where E_i is computed:

for $i = 1, 2, \ldots, 8$

$$\begin{aligned} E_i &= 0 \text{ if } V_i < V_0 \\ E_i &= 1 \text{ if } V_i = V_0, \\ E_i &= 2 \text{ if } V_i > V_0 \end{aligned} \tag{8.92}$$

and the element E_i occupies the same position as pixel i. Because each element of TU has one of three possible values, the combination of all eight elements results in $3^8 = 6{,}561$ possible texture units. There is no unique way to label and order the 6,561 texture units. Therefore, the texture unit of a 3 × 3 neighborhood of pixels (Figure 8-57b–d) is computed:

$$N_{\mathrm{TU}} = \sum_{i=1}^{8} 3^{i-1} E_i \tag{8.93}$$

where E_i is the ith element of the texture unit set $\mathrm{TU} = \{E_1, E_2, \ldots, E_8\}$. The first element, E_i, may take any one of the eight possible positions from *a* through *h* in Figure 8-57a. An example of transforming a 3 × 3 neighborhood of image brightness values into a texture unit (TU) and a texture unit number (N_{TU}) using the ordering method starting at *a* is shown in Figure 8-57. In this example, the texture unit number, N_{TU}, for the central pixel has a value of 6,095. The eight brightness values in the hypothetical neighborhood are very diverse (that is, there is a lot of heterogeneity in this small region of the image); therefore, it is not surprising that the central pixel has such a high texture unit number. Eight separate texture unit numbers could be calculated for this central pixel based on the eight ways of ordering shown in Figure 8-57a. The eight N_{TU} values could then be averaged to obtain a mean N_{TU} value for the central pixel.

The possible texture unit values range from 0 to 6,560 and describe the local texture of a pixel in relationship to its eight neighbors. The frequency of occurrence of all the pixel texture unit numbers over a whole image is called the *texture spectrum*. It may be viewed in a graph with the range of possible texture unit numbers (N_{TU}) on the x-axis (values from 0 to 6,560) and the frequency of occurrence on the y-axis (Figure 8-58). Each image (or subimage) should have a unique texture spectrum if its texture is truly different from other images (or subimages).

Wang and He (1990) developed algorithms for extracting textural features from the texture spectrum of an

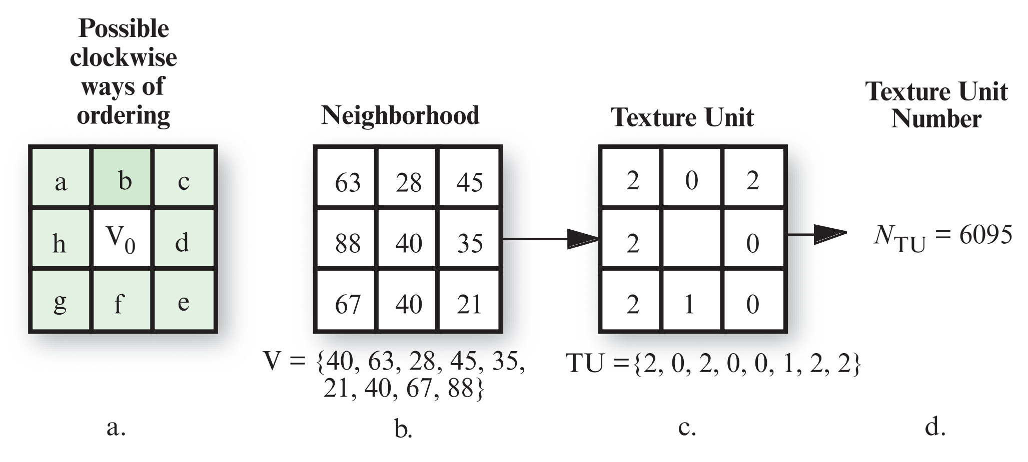


FIGURE 8–57 How a 3 × 3 neighborhood of brightness values is transformed into a texture unit number (N_{TU}), which has values ranging from 0 to 6,560. a) Possible clockwise ways of ordering the eight elements of the texture unit. The first element E_i in Equation 8.93 may take any of the eight positions from *a* through *h*. In this example, the ordering position begins at *a*. b) Brightness values found in the 3 × 3 neighborhood. This is a very heterogeneous group of pixels and should result in a high texture unit number. c) Transformation of the neighborhood brightness values into a texture unit. d) Computation of the texture unit number based on Equation 8.93 (values range from 0 to 6,560). It is possible to compute eight separate texture unit numbers from this neighborhood and then take the mean (based on Wang and He, 1990).

image, including black–white symmetry, geometric symmetry, and degree of direction. Only the geometric symmetry measure is presented here. For a given texture spectrum, let $S_j(i)$ be the occurrence frequency of the texture unit numbered i in the texture spectrum under the ordering way j, where $i = 0, 1, 2, \ldots, 6560$ and $j = 1, 2, 3, \ldots, 8$ (the ordering ways *a, b, c, …, h* are, respectively, represented by $j = 1, 2, 3, \ldots, 8$). Geometric symmetry (GS) for a given image (or subimage) is:

$$GS = \left[1 - \frac{1}{4}\sum_{j=1}^{4}\frac{\sum_{i=0}^{6560}\left|S_j(i) - S_{j+4}(i)\right|}{2 \times \sum_{i=0}^{6560} S_j(i)}\right] \times 100\,. \qquad (8.94)$$

GS values are normalized from 0 to 100 and measure the symmetry between the spectra under the ordering ways *a* and *e*, *b* and *f*, *c* and *g*, and *d* and *h* for a given image. This measure provides information on the shape regularity of images (Wang and He, 1990). A high value means the texture spectrum will remain approximately the same even if the image is rotated 180°. The degree of direction measure provides information about the orientation characteristics of images.

Texture Statistics Based on the Semi-variogram

Numerous authors have investigated the use of the semi-variogram discussed in Chapter 4 to derive texture information (e.g., Woodcock et al., 1988; Lark, 1996; Rodriguez-Galiano et al., 2012). Maillard (2003) identified the problems associated with using different

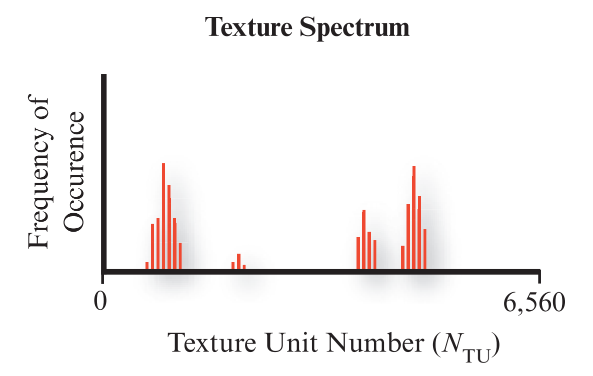


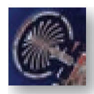
FIGURE 8–58 Hypothetical texture spectrum derived from an image or subimage.

variogram models (e.g., spherical, exponential, and sinusoidal) and other criteria. He developed a variogram texture operator that:

- uses a rather large window to cover larger distance lags (up to 32 pixels),
- is rotation-invariant and preserves anisotrophy, and
- incorporates the mean square-root pair difference function (SRPD) as a semi-variance estimator.

Maillard compared the variogram, GLCM, and Fourier-based texture measures and concluded that the variogram and GLCM texture measures were generally superior. Rodriguez-Galiano et al. (2012) used five different types of geostatistical textural measures for three different window sizes and three different lags yielding a total of 972 potential input variables. They found the geostatistical texture measures to be more useful than the GLCM texture measures.

Although texture features have been increasingly incorporated into multispectral classifications, no single algorithm that combines efficiency and effectiveness has yet to be widely adopted. Also, the texture features derived for one type of application (e.g., land-use classification at the urban fringe) are not necessarily useful when applied to another geographic problem (e.g., identification of selected geomorphic classes). Finally, some parameters central to the computation of the texture features are still derived empirically (e.g., the size of the window or the location of certain thresholds). This makes it difficult to compare and contrast studies when so many variables in the creation of the texture features are not held constant.



Landscape Ecology Metrics

Vegetation indices are useful for monitoring the condition and health of vegetated pixels. However, such per-pixel analysis does not provide any information about the nature of surrounding pixels, including their site and association characteristics. Landscape ecology principles have been developed that increasingly incorporate remote sensor data to assess the health and diversity of vegetation and other variables within entire ecosystems. This has resulted in the development of numerous **landscape ecology metrics** or **indicators** that are of significant value when analyzing rangeland, grassland, forests, and wetland (Frohn, 1998; Frohn and Hao, 2006; McGarigal et al., 2013). Numerous government agencies, such as the Environmental Protection Agency, base much of their environmental modeling and landscape characterization on these metrics and indicators (EPA Landscape Ecology, 2014). Therefore, it is useful to provide a brief review of their origin and summarize several of the more important landscape ecology metric variables that can be extracted from remotely sensed data.

The term *landscape ecology* was first introduced by the German geographer Carl Troll (1939) who made widespread use of the then new technique of aerial photography. Troll intended for the term landscape ecology to distinguish his approach for using such imagery to interpret the interaction of water, land surfaces, soil, vegetation, and land use from that of conventional photographic interpretation and cartography. Landscape ecology has been intensively practiced in Europe for many decades and became generally recognized in the United States in about 1980. Since then, landscape ecology has rapidly evolved as a discipline, spurred by the synergistic interactions between remote sensing and GIS techniques and advances in ecological theory.

Landscape ecology is the study of the structure, function, and changes in heterogeneous land areas composed of interacting organisms (Bourgeron and Jensen, 1993). It is the study of the interaction between landscape patterns and ecological processes, especially the influence of landscape pattern on the flows of water, energy, nutrients, and biota. What distinguishes landscape ecology from the many separate disciplines that it embraces (e.g., geography, biology, ecology, hydrology) is that it provides a hierarchical framework for interpreting ecological structure, function, change, and resiliency at multiple scales of inquiry.

Traditional measures to protect the environment, such as preventing water pollution or protecting biodiversity, often focused on specific effluent discharges or fine-scale habitat requirements. This method has been described as the fine-filter approach. In contrast, the coarse-filter approach to resource conservation states that "by managing aggregates (e.g., communities, ecosystems, landscapes), the components of these aggregates will be managed as well" (Bourgeron and Jensen, 1993). In other words, the most cost-effective strategy to maintain the resiliency and productivity of ecological systems is to conserve (or restore) the diversity of species, ecosystem processes, and landscape patterns that create the systems. Applying this coarse-filter management method requires that landscape patterns be evaluated at multiple spatial and temporal scales rather than simply at the traditional scales of stream reach or forest stand.

Hierarchy theory allows landscape ecologists to integrate multiple scales of information to determine whether landscape patterns are sufficient to allow ecological processes to operate at the necessary scales. The objective is to investigate changes in the distribution, dominance, and connectivity of ecosystem components and the effect of these changes on ecological and biological resources. For example, ecosystem fragmentation has been implicated in the decline of biological diversity and ecosystem sustainability at a number of spatial scales. Determining status and trends in the pattern of landscapes is critical to understanding the overall condition of ecological resources. Landscape patterns thus provide a set of indicators (e.g., pattern shape, dominance, connectivity, configuration) that can be used to assess ecological status and trends at a variety of scales.

A hierarchical framework also permits two important types of comparisons: 1) to compare conditions within and across landscapes, and 2) to compare conditions across different types of ecological risks. Such ecological risks include the risk of erosion, loss of soil productivity, loss of hydrologic function, and loss of biodiversity.

TABLE 8–11 Societal values, example indicators, and candidate landscape ecology metrics.

Societal Value	Indicator	Candidate Metrics
Biodiversity	wildlife habitat suitability	patch statistics (number, total area, average size, largest size, distance between, ratio perimeter to area, shape, fractal dimension, square pixel model, etc.), fragmentation, contagion, zone fragmentation index, patch per-unit-area index, dominance, adjacency of land-cover types, Shannon diversity, biophysical attribute patterns
	stream biological condition	diversity, square pixel model, dominance, fragmentation, zone fragmentation index, patch per-unit-area index, adjacency of land-cover types, slope, elevation, diffusion rates, percolation threshold, erosion index, texture, biophysical attribute patterns, geochemical attributes
	forest plant species richness	diversity, dominance, fragmentation, zone fragmentation index, patch per-unit-area index, slope, erosion index, texture, patch statistics, square pixel model, biophysical attribute patterns
	landscape sustainability	patch statistics, contagion, zone fragmentation index, patch per-unit-area index, fragmentation, texture, dominance, fractal dimension, square pixel model, biophysical attribute patterns
Watershed integrity	water quality	patch statistics, erosion index, hydrologic modification, adjacency of land-cover types, dominance, contagion, zone fragmentation index, patch per unit area index, fractal dimension, square pixel model, elevation, slope, biophysical attribute patterns, geochemical attributes
	vulnerability to flooding	patch statistics, adjacency of land-cover types, erosion index, dominance, contagion, zone fragmentation index, patch per-unit-area index, fractal dimension, square pixel model, hydrologic modification, elevation, slope, texture, biophysical attribute patterns
Landscape resilience	landscape sustainability	patch statistics, dominance, contagion, zone fragmentation index, patch per-unit-area index, fragmentation, fractal dimension, square pixel model, biophysical attribute patterns

Scalable units are needed to address landscape ecology issues at multiple scales within a hierarchical framework. Examples of scalable units include patches, patterns, and landscapes. A *patch unit* is a set of contiguous measurement units (e.g., pixels) that have the same numerical value. A *pattern unit* is a collection of measurement units and/or patch units that have the property of being the minimum unit descriptor of a larger spatial area. The scales of assessment questions and indicators suggest two types of landscape units: watersheds and landscape pattern types (LPT) (Wickham and Norton, 1994). Watersheds and LPTs capture or bound four important flow processes operating within and among landscapes: flows of energy, water, nutrients, and biota. Scales of watersheds and LPTs range from approximately 10^3 to 10^6 units in extent, and from 1 to 100 ha.

Landscape Indicators and Patch Metrics

Jones et al. (1998) suggest that landscape integrity can be monitored by carefully watching the status of the following indicators:

- land-cover composition and pattern
- riparian extent and distribution
- ground water
- greenness pattern
- degree of biophysical constraints
- erosion potential.

Monitoring these landscape indicators requires precise, repeatable measurements of terrain *patches* such as individual forest stands, rangeland, wetland, and/or agricultural fields (Table 8-11). It is also important to identify patches of pure urban structure such as residential and commercial land use. These measurements of terrain patches are routinely referred to as *landscape pattern* and *structure metrics* (McGarigal et al., 2013). Numerous landscape structure metrics have been developed (e.g., Weiers et al., 2004; Frohn and Hao, 2006). Many of these landscape ecology metrics are available in *FRAGSTATS—A Spatial Pattern Analysis Program for Categorical Maps* (McGarigal et al., 2013). Scientists often use this program to extract patch, class, and landscape indices.

O'Neill et al. (1997) suggest that the health of an ecosystem could be monitored if the following three landscape ecology metrics were monitored through time: dominance, contagion, and fractal dimension.

Dominance, *D*, is the information theoretic index that identifies the extent to which the landscape is dominated by a single land-cover type. The metric, $0 < D < 1$, is computed as

$$D = 1 - \left[\sum_{k} \frac{(-P_k \times \ln P_k)}{\ln(n)} \right] \tag{8.95}$$

where $0 < P_k < 1$ is the proportion of land-cover type k, and n is the total number of land-cover types present in the landscape.

Contagion, C, expresses the probability that land cover is more "clumped" than the random expectation. The index, $0 < C < 1$, is

$$C = 1 - \left[\sum_{i} \sum_{j} \frac{(-P_{ij} \times \ln P_{ij})}{2\ln(n)} \right] \tag{8.96}$$

where P_{ij} is the probability that a pixel of cover type i is adjacent to type j.

The *fractal dimension*, *F,* of patches indicates the extent of human reshaping of landscape structure (O'Neill et al., 1997). Humans create simple landscape patterns; nature creates complex patterns. The fractal dimension index is calculated by regressing the log of the patch perimeter against the log of the patch area for each patch on the landscape. The index equals twice the slope of the regression line. Patches of four or fewer pixels are excluded because resolution problems distort their true shape.

O'Neill et al. (1997) suggest that this set of three indices may capture fundamental aspects of landscape pattern that influence ecological processes. Significant changes in these indices for an ecosystem might indicate that perhaps deleterious processes are at work in the environment. For example, consider a small ecosystem that exhibits a less modified fractal dimension, is highly clumped, and has relatively few land-cover types within it. It might appear in a three-dimensional landscape metric space at location *a* in Figure 8-59. If this small ecosystem were subdivided with several new roads and fragmented, its location might move in three-space to *b* with many land-cover types being introduced (dominance change), be less clumped (contagion), with a more modified fractal dimension. This could be good or bad. In fact, the relationships between the metric values and how they actually relate to ecological principles are still being determined. Remote sensing of vegetation within these patches is very important and constitutes one of the major factors responsible for whether or not the metrics are robust and useful for ecological modeling.

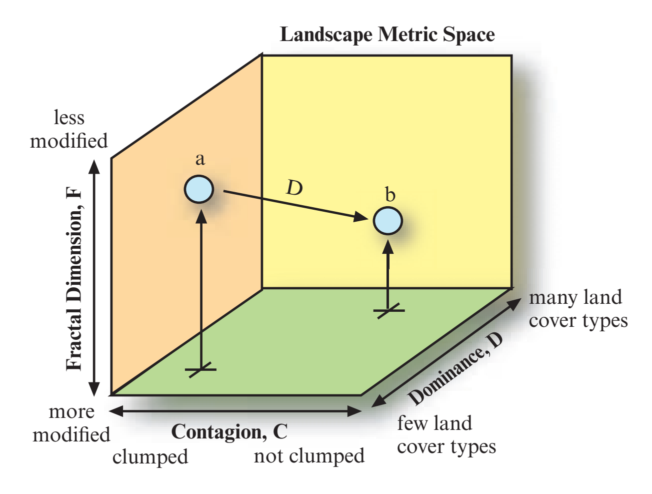


FIGURE 8–59 A hypothetical three-dimensional landscape metric feature space. Point *a* represents where a stable, unmodified ecosystem might reside in dominance, contagion, and fractal dimension feature space. When the landscape is modified by man or natural forces, it may move to point *b* with many land-cover types being introduced (dominance change), be less clumped (contagion), and exhibit a more modified fractal dimension. This could be good or bad. Remote sensing of vegetation is critical to the extraction of landscape patch metrics.

Ritters et al. (1995) reviewed 55 patch metrics and concluded that the following metrics accounted for most of the variance, including: the number of attribute cover types in a region (n); contagion (previously discussed); average perimeter-area ratio; perimeter-area scaling; average large-patch density-area scaling; standardized patch shape; and patch perimeter-area scaling. The algorithms for all 55 metrics are provided in the paper.

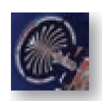
Applying the principles of landscape ecology requires an understanding of the natural variability of landscape patterns and processes across both space and time. Estimates of this variability are essential to determining whether the current condition of landscape is sustainable, given its historic patterns and processes. Moreover, the estimates are extremely useful in both broad-level assessment of risk to resources, as well as to finer-scale assessments.

The goal is usually to

- estimate, on a regional basis and with known confidence, the current status, trends, and changes in selected *indicators* of the important landscapes;
- estimate with known confidence the geographic coverage and extent of the landscapes' patterns and types;

- seek associations between selected indicators of natural and anthropogenic stressors and indicators of landscape condition; and
- provide statistical summaries and periodic assessments of the condition of the landscapes.

Monitoring the vegetation in landscape patches using remotely sensed data and deriving accurate metrics from these data is a very important component of many ecosystem monitoring programs. For example, Arroyo-Moya et al. (2005) used remote sensing-derived landscape ecology metrics to analyze forest fragmentation and regrowth characteristics in Costa Rica from 1960–2000. Jackson and Jensen (2005) used remote sensing-derived landscape ecology metrics to perform reservoir shoreline environmental monitoring in South Carolina. Frohn and Hao (2006) provide an overview of numerous landscape ecology metrics applied to their study of deforestation in Rondonia, Brazil.



References

Ali Baig, Zhang, L., Shuai, T., and Q. Tong, 2014, "Derivation of a Tasselled Cap Transformation based on Landsat 8 At-Satellite Reflectance," *Remote Sensing Letters*, 5(5):423–431.

Amato, U., Antoniadis, A., Cuomo, V., Cutillo, L., Franzese, M., Murino, L., and C. Serio, 2008, "Statistical Cloud Detection from SEVIRI Multispectral Images," *Remote Sensing of Environment*, 112:750–766.

Arroyo-Moya, J. P., Sanchez-Azofeifa, G. A., Rivard, B., Calvo, J. C., and D. H. Janzen, 2005, "Dynamics in Landscape Structure and Composition for Chorotega Region, Costa Rica from 1960 to 2000," *Agriculture, Ecosystems and Environment*, 106:27–39.

Baranoski, G. V. G., and J. G. Rokne, 2005, "A Practical Approach for Estimating the Red Edge Position of Plant Leaf Reflectance," *International Journal of Remote Sensing*, 26(3):503–521.

Birth, G. S., and G. McVey, 1968, "Measuring the Color of Growing Turf with a Reflectance Spectroradiometer," *Agronomy Journal*, 60:640–643.

Boer, M. M., Macfarlane, C., Norris, J., Sadler, R. J., Wallace, J., and P. F. Grierson, 2008, "Mapping Burned Areas and Burn Severity Patterns in SW Australia using Remotely-sensed Changes in Leaf Area Index," *Remote Sensing of Environment*, 112:4358–4369.

Bourgeron, P. S., and M. E. Jensen, 1993, "An Overview of Ecological Principles for Ecosystem Management," in Jensen, M., and P. Bourgeron (Eds.), *Eastside Forest Ecosystem Health Assessment, Vol. II, Ecosystem Management: Principles and Applications*, Missoula: USDA Forest Service, 49–60.

Brewer, C. K., Winne, J. C., Redmond, R. L., Optiz, D. W., and M. V. Mangrich, 2005, "Classifying and Mapping Wildfire Sevirity: A Comparison of Methods," *Photogrammetric Engineering & Remote Sensing*, 71:1311–1320.

Briggs, J. M., and M. D. Nellis, 1991, "Seasonal Variation of Heterogeneity in the Tallgrass Prairie: A Quantitative Measure Using Remote Sensing," *Photogrammetric Engineering & Remote Sensing*, 57(4):407–411.

Broge, N. H., and E. Leblanc, 2000, "Comparing Prediction Power and Stability of Broadband and Hyperspectral Vegetation Indices for Estimation of Green Leaf Area Index and Canopy Chlorophyll Density," *Remote Sensing of Environment*, 76:156–172.

Carr, J. R., 1998, "A Visual Basic Program for Principal Components Transformation of Digital Images," *Computers & Geosciences,* 24(3):209–281.

Carter, G. A., 1991, "Primary and Secondary Effects of the Water Content on the Spectral Reflectance of Leaves," *American Journal of Botany,* 78(7):916–924.

Carter, G. A., 1993, "Responses of Leaf Spectral Reflectance to Plant Stress," *American J. of Botany*, 80(3):231–243.

Carter, G. A., Cibula, W. G., and R. L. Miller, 1996, "Narrow-band Reflectance Imagery Compared with Thermal Imagery for Early Detection of Plant Stress," *Journal of Plant Physiology*, 148:515–522.

Chavez, P. C., Guptill, S. C., and J. A. Bowell, 1984, "Image Processing Techniques for Thematic Mapper Data," *Proceedings*, Annual Meeting of the American Society for Photogrammetry & Remote Sensing, 2:728–743.

Chen, J. M., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S. G., White, H. P., Hall, R. J., Peddle, D. R., King, D. J., Trofymow, J. A., Swift, E., Van der Sanden, J., and P. K. Pellikka, 2002, "Derivation and Validation of Canada-wide Coarse-resolution Leaf Area Index Maps Using High-resolution Satellite Imagery and Ground Measurements," *Remote Sensing of Environment*, 80:165–184.

Cheng, Y., Ustin, S., Riano, D., and V. Vanderbilt, 2008, "Water Content Estimation from Hyperspectral Images and MODIS Indexes in Southeastern Arizona," *Remote Sensing of Environment*, 112:363–374.

Cibula, W. G., and G. A. Carter, 1992, "Identification of a Far-Red Reflectance Response to *Ectomycorrhizae* in Slash Pine," *International Journal of Remote Sensing*, 13(5):925–932.

Clausi, D. A., 2002, "An Analysis of Co-occurrence Texture Statistics as a Function of Grey Level Quantization," *Canadian Journal of Remote Sensing*, 28(1):45–62.

Clevers, J. G., 1994, "Imaging Spectrometry in Agriculture: Plant Vitality and Yield Indicators," in Hill, J. and J. Megier (Eds.), *Imaging Spectrometry: A Tool for Environmental*

Observations, Dordrecht, Netherlands: Kluwer Academic, 193–219.

Cohen, W. B., 1991, "Response of Vegetation Indices to Changes in Three Measures of Leaf Water Stress," *Photogrammetric Engineering & Remote Sensing*, 57(2):195–202.

Collins, W., 1978, "Remote Sensing of Crop Type and Maturity," *Photogrammetric Engineering & Remote Sensing*, 44(1):43–55.

Colombo, R., Bellingeri, D., Fasolini, D., and C. M. Marino, 2003, "Retrieval of Leaf Area Index in Different Vegetation Types Using High Resolution Satellite Data," *Remote Sensing of Environment*, 86:120–131.

Crippen, R. E., 1988, "The Dangers of Underestimating the Importance of Data Adjustments in Band Ratioing," *International Journal of Remote Sensing*, 9(4):767–776.

Crippen, R. E., and R. G. Blom, 2001, "Unveiling the Lithology of Vegetated Terrains in Remotely Sensed Imagery," *Photogrammetric Engineering & Remote Sensing*, 67(8):935–943.

Crist, E. P., and R. C. Cicone, 1984, "Application of the Tasseled Cap Concept to Simulated Thematic Mapper Data," *Photogrammetric Engineering & Remote Sensing*, 50:343–352.

Crist, E. P., and R. J. Kauth, 1986, "The Tasseled Cap Demystified," *Photogrammetric Engineering & Remote Sensing*, 52(1):81–86.

Culbert, P. D., with 6 co-authors, 2012, "Modeling Broadscale Patterns of Avian Species Richness Across the Midwestern United States with Measures of Satellite Image Texture," *Remote Sensing of Environment*, 118:140–150.

Dash, J., Jeganathan, C., and P. M. Atkinson, 2010, "The Use of MERIS Terrestrial Chlorophyll Index to Study Spatial-Temporal Variation in Vegetation Phenology over India," *Remote Sensing of Environment*, 114:1388–1402.

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., Brown de Colstoun, E., and J. E. McMurtrey III, 2000, "Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance," *Remote Sensing of Environment*, 74:229–239.

Dawson, T. P., and P. J. Curran, 1998, "A New Technique for Interpolating the Reflectance Red Edge Position," *International Journal of Remote Sensing*, 19(11):2133–2139.

Deering, D. W., Rouse, J. W., Haas, R. H., and J. A. Schell, 1975, "Measuring Forage Production of Grazing Units from Landsat MSS Data," *Proceedings, 10th International Symposium on Remote Sensing of Environment*, 2:1169–1178.

Didan, K., 2002, *MODIS Vegetation Index Production Algorithms*, MODIS Vegetation Workshop, Missoula, Montana, July 15; Terrestrial Biophysics and Remote Sensing (TBRS) MODIS Team, Tucson: University of Arizona, www.ntsg.umt.edu/MODISCon/index.html.

Dronova, I., Gong, P., Clinton, N., Wang, L., Fu, A., Qi, S., and Y. Liu, 2012, "Landscape Analysis of Wetland Plant Functional Types: The Effects of Image Segmentation Scale, Vegetation Classes and Classification Methods," *Remote Sensing of Environment*, 127:357–369.

Du, Y., Teillet, P. M., and J. Cihlar, 2002, "Radiometric Normalization of Multitemporal High-resolution Satellite Images with Quality Control for Land Cover Change Detection," *Remote Sensing of Environment*, 82:123–134.

Eastman, J. R., and M. Fulk, 1993, "Long Sequence Time Series Evaluation Using Standardized Principal Components," *Photogrammetric Engineering & Remote Sensing*, 59(6):991–996.

Eliason, E. M., and A. S. McEwen, 1990, "Adaptive Box Filters for Removal of Random Noise from Digital Images," *Photogrammetric Engineering & Remote Sensing*, 56(4):453–458.

ENVI Suppression, 2013, *ENVI EX User's Guide*, Boulder: ITT Visual Information Solutions, 90–91.

EPA Landscape Ecology, 2014, *Selected Landscape Ecology Projects*, Washington: EPA, www.epa.gov/nerlesd1/land-sci/default.htm.

ERDAS, 2013, *ERDAS Field Guide*, Atlanta: Intergraph, Inc., 772 p.

Estes, L. D., Okin, G. S., Mwangi, A. G., and H. H. Shugart, 2008, "Habitat Select by Rare Forest Antelope: A Multiscale Approach Combining Field Data and Imagery from Three Sensors," *Remote Sensing of Environmen*t, 112:2033–2050.

Estes, L. D., Reillo, P. R., Mwangi, A. G., Okin, G. S., and H. H. Shugart, 2010, "Remote Sensing of Structural Complexity Indices for Habitat and Species Distribution Modeling," *Remote Sensing of Environment*, 114:792–804.

Ferro, C. J., and T. A. Warner, 2002, "Scale and Texture Digital Image Classification," *Photogrammetric Engineering & Remote Sensing*, 68(1):51–63.

Frohn, R. C., 1998, *Remote Sensing for Landscape Ecology*, Boca Raton: Lewis, 99 p.

Frohn, R. C., and Y. Hao, 2006, "Landscape Metric Performance in Analyzing Two Decades of Deforestation in the Amazon Basin of Rondonia, Brazil," *Remote Sensing of Environment*, 100:237–251.

Galvao, L. S., Formaggio, A. R., and D. A. Tisot, 2005, "Discrimination of Sugarcane Varieties in Southeastern Brazil with EO-1 Hyperion Data," *Remote Sensing of Environment*, 94(4),523–534.

Galvao, L. S., Roberts, D. A., Formaggio, A., Numata, I., and F. Breunig, 2009, "View Angle Effects on the Discrimination of Soybean Varieties and on the Relationships between Vegetation Indices and Yield using Off-nadir Hyperion Data," *Remote Sensing of Environment*, 113:846–856.

Gao, B., 1996, "NDWI–A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space," *Remote Sensing of Environment*, 58_257–266.

Gates, D. M., Keegan, J. J., Schleter, J. C., and V. R. Weidner, 1965, "Spectral Properties of Plants," *Applied Optics*, 4(1):11–20.

Gausmann, H. W., Allen, W. A., and R. Cardenas, 1969, "Reflectance of Cotton Leaves and their Structure," *Remote Sensing of Environment*, 1:110–122.

Gil, B., Mitiche, A., and J. K. Aggarwal, 1983, "Experiments in Combining Intensity and Range Edge Maps," *Computer Vision, Graphics, and Image Processing*, 21:395–411.

Gitelson, A. A., Kaufman, Y. J., Stark, R., and D. Rundquist, 2002, "Novel Algorithms for Remote Estimation of Vegetation Fraction," *Remote Sensing of Environment*, 80:76–87.

Gong, P., Marceau, D. J., and P. J. Howarth, 1992, "A Comparison of Spatial Feature Extraction Algorithms for Land-Use Classification with SPOT HRV Data," *Remote Sensing of Environment,* 40:137–151.

Gonzalez, R. C., and R. E. Woods, 2007, *Digital Image Processing*. 3rd Ed., Reading, MA: Addison-Wesley, 976 p.

Gonzalez, R. C., and P. Wintz, 1977, *Digital Image Processing*. Reading: Addison-Wesley, 431 p.

Good, E. J., Kong, X., Embury, O., Merchant, C. J., and J. J. Remedios, 2012, "An Infrared Desert Dust Index for the Along-Track Scanning Radiometers," *Remote Sensing of Environment*, 116:159–176.

Goward, S. N., Markham, B., Dye, D. G., Dulaney, W., and J. Yang, 1991, "Normalized Difference Vegetation Index Measurements from the AVHRR," *Remote Sensing of Environment*, 35:257–277.

Gray, J., and C. Song, 2012, "Mapping Leaf Area Index using Spatial, Spectral, and Temporal Information from Multiple Sensors," *Remote Sensing of Environment*, 119:173–183.

Guerschman, J. P., Hill, M. J., Renzullo, L. J., Barrett, D. J., Marks, A. S., and E. J. Botha, 2009, "Estimating Fractional Cover of Photosynthetic Vegetation, Non-photosynthetic Vegetation and Bare Soil in the Australian Tropical Savanna Region Upscaling the Eo-1 Hyperion and MODIS Sensors," *Remote Sensing of Environment*, 113:928–945.

Guyon, D., Guillot, M., Vitasse, Y., Cardot, H., Hagolle, O., Delszon, S., and J. Wigneron, 2011, "Monitoring Elevation Variations in Leaf Phenology of Deciduous Broadleaf Forests from SPOT/VEGETATION Time-series," *Remote Sensing of Environment*, 115:615–627.

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., and L. Dextraze, 2002, "Integrated Narrow-band Vegetation Indices for Prediction of Crop Chlorophyll Content for Application to Precision Agriculture," *Remote Sensing of Environment*, 81:416–426.

Hall-Beyer, M., 2007, *GLCM Tutorial Home Page,* Version 2.1, http://www.fp.ucalgary.ca/mhallbey/tutorial.htm.

Haralick, R. M., 1979, "Statistical and Structural Approaches to Texture," *Proceedings of the IEEE*, 67:786–804.

Haralick, R. M., 1986, "Statistical Image Texture Analysis," T. Y. Young and K. S. Fu (Eds.), *Handbook of Pattern Recognition and Image Processing*, New York: Academic Press, 247–280.

Haralick, R. M., and K. Fu, 1983, "Pattern Recognition and Classification," Chapter 18 in R. N. Colwell (Ed.), *Manual of Remote Sensing*, Falls Church, VA: American Society of Photogrammetry, 793–805.

Haralick, R. M., and K. S. Shanmugam, 1974, "Combined Spectral and Spatial Processing of ERTS Imagery Data," *Remote Sensing of Environment*, 3:3–13.

Haralick, R. M., Shanmugan, K., and I. Dinstein, 1973, "Texture Feature for Image Classification," *IEEE Transactions Systems, Man and Cybernetic*s, SMC-3:610–621.

Hardisky, M. A., Klemas, V., and R. M. Smart, 1983, "The Influence of Soil Salinity, Growth Form, and Leaf Moisture on the Spectral Radiance of *Spartina alterniflora* Canopies," *Photogrammetric Engineering & Remote Sensing*, 49(1):77–83.

He, D. C., and L. Wang, 1990, "Texture Unit, Texture Spectrum, and Texture Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):509–512.

Herold, M., Liu, X., Hang, L., and K. C. Clarke, 2003, "Spatial Metrics and Image Texture for Mapping Urban Land Use," *Photogrammetric Engineering & Remote Sensing*, 69(9):991–1001.

Houborg, R., Soegaard, H., and E. Boegh, 2007, "Combining Vegetation Index and Model Inversion Methods for the Extraction of Key Vegetation Biophysical Parameters using Terra and Aqua MODIS Reflectance Data," *Remote Sensing of Environment*, 106:39–58.

Huang, C., Wylie, B., Yang, L., Homer, C., and G. Zylstra, 2002, "Derivation of a Tasseled Cap Transformation based on Landsat 7 at-Satellite Reflectance," *International Journal of Remote Sensing*, 23(8):1741–1748.

Huete, A. R., 1988, "A Soil-adjusted Vegetation Index (SAVI)," *Remote Sensing of Environment*, 25:295–309.

Huete, A. R., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and G. Ferreira, 2002a, "Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices," *Remote Sensing of Environment*, 83:195–213.

Huete, A. R., Didan, K., and Y. Yin, 2002b, *MODIS Vegetation Workshop*, Missoula, Montana, July 15–18; Terrestrial Biophysics and Remote Sensing (TBRS) MODIS Team, University of Arizona.

Huete, A. R., Hua, G., Qi, J., Chehbouni A., and W. J. van Leeuwem, 1992, "Normalization of Multidirectional Red and Near-infrared Reflectances with the SAVI," *Remote Sensing of Environment*, 40:1–20.

Huete, A. R., and C. Justice, 1999, *MODIS Vegetation Index (MOD 13) Algorithm Theoretical Basis Document*, Greenbelt: NASA GSFC, http://modarch.gsfc.nasa.gov /MODIS/LAND/#vegetation-indices, 129 p.

Huete, A. R., and H. Q. Liu, 1994, "An Error and Sensitivity Analysis of the Atmospheric and Soil-Correcting Variants of the NDVI for the MODIS-EOS," *IEEE Transactions on Geoscience and Remote Sensing*, 32(4):897–905.

Huete, A. R., Liu, H. Q., Batchily, K., and W. J. van Leeuwen, 1997, "A Comparison of Vegetation Indices Over a Global Set of TM Images for EOS-MODIS," *Remote Sensing of Environment*, 59:440–451.

Hunt, E. R., Rock, B. N., and P. S. Nobel, 1987, "Measurement of Leaf Relative Water Content by Infrared Reflectance," *Remote Sensing of Environment*, 22:429–435.

Im, J., and J. R. Jensen, 2009, "Hyperspectral Remote Sensing of Vegetation," *Geography Compass*, Vol. 3 (November), DOI: 10.1111/j.1749-8198.2008.00182.x.

Jackson, R. D., 1983, "Spectral Indices in *n*-Space," *Remote Sensing of Environment*, 13:409–421.

Jackson, M., and J. R. Jensen, 2005, "Evaluation of Remote Sensing-derived Landscape Ecology Metrics for Reservoir Shoreline Environmental Monitoring," *Photogrammetric Engineering & Remote Sensing*, 71(12):1387–1397.

Jackson, T. J. and seven co-authors, 2004, "Vegetation Water Content Mapping using Landsat Data Derived Normalized Difference Water Index for Corn and Soybeans," *Remote Sensing of Environment*, 92:475–482.

Jahne, B., 2005, *Digital Image Processing*, NY: Springer, 607 p.

Jain, A. K., 1989, *Fundamentals of Digital Image Processing*, Englewood Cliffs, NJ: Prentice-Hall, 342–357.

Jensen, J. R., Hadley, B. C., Tullis, J. A., Gladden, J., Nelson, E., Riley, S., Filippi, T., and M. Pendergast, 2003, *Hyperspectral Analysis of Hazardous Waste Sites on the Savannah River Site in 2002*, Westinghouse Savannah River Company: Aiken, WSRC-TR-2003-00275, 52 p.

Jensen, J. R., Hodgson, M. E., Christensen, E., Mackey, H. E., Tinney, L. R., and R. Sharitz, 1986, "Remote Sensing Inland Wetlands: A Multispectral Approach," *Photogrammetric Engineering & Remote Sensing*, 52(2):87–100.

Jensen, J. R., Im, J., Hardin, P., and R. R. Jensen, 2009, "Chapter 19: Image Classification," in *The Sage Handbook of Remote Sensing*, Warner, T. A., Nellis, M. D. and G. M. Foody, (Eds.), 269–296.

Jensen, J. R., Pace, P. J., and E. J. Christensen, 1983, "Remote Sensing Temperature Mapping: The Thermal Plume Example," *American Cartographer*, 10:111–127.

Jensen, J. R., Lin, H., Yang, X., Ramsey, E., Davis, B., and C. Thoemke, 1991, "Measurement of Mangrove Characteristics in Southwest Florida Using SPOT Multispectral Data," *Geocarto International*, 2:13–21.

Jiang, Z., Huete, A. R., Didan, K., and T. Miura, 2008, "Development of A Two-band Enhanced Vegetation Index without a Blue Band," *Remote Sensing of Environment*, 112:3833–3845.

Jin, S., and S. A. Sader, 2005, "Comparison of Time Series Tasseled Cap Wetness and the Normalized Difference Moisture Index in Detecting Forest Disturbances," *Remote Sensing of Environment*, 94:364–372.

Jones, K. B., Ritters, K. H., Wickham, J. D., Tankersley, R. D., O'Neill, R. V., Chaloud, D. J., Smith, E. R., and A. C. Neale, 1998, *Ecological Assessment of the United States: Mid-Atlantic Region*, Washington: EPA, 103 p.

Karnieli, A., Kaufman, Y. J., Remer, L., and A. Wald, 2001, "AFRI: Aerosol Free Vegetation Index," *Remote Sensing of Environment*, 77:10–21.

Kaufman, Y. J., Karnieli, A., and D. Tanre, 2000, "Detection of Dust Over Deserts Using Satellite Data in the Solar Wavelengths," *IEEE Transactions on Geoscience and Remote Sensing*, 38:525–531.

Kaufman, Y. J., and D. Tanre, 1992, "Atmospherically Resistant Vegetation Index (ARVI) for EOS-MODIS," *IEEE Transactions on Geoscience and Remote Sensing*, 30(2):261–270.

Kauth, R. J., and G. S. Thomas, 1976, "The Tasseled Cap—A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat," *Proceedings, Symposium on Machine Processing of Remotely Sensed Data*, West Lafayette, IN: LARS, 41–51.

Kauth, R. J., Lambeck, P. F., Richardson, W., Thomas, G.S., and A. P. Pentland, 1979, "Feature Extraction Applied to Agricultural Crops as Seen by Landsat," *Proceedings, LACIE Symposium,* Houston: NASA, 705–721.

Kayitakire, F., Hamel, C., and P. Defourny, 2006, "Retrieving Forest Structure Variables Based on Image Texture Analysis and IKONOS-2 Imagery," *Remote Sensing of Environment,* 102:390–401.

Kim, M., Daughtry, C. S., Chappelle, E. W., McMurtrey III, J. E., and C. L. Walthall, 1994, "The Use of High Spectral Resolution Bands for Estimating Absorbed Photosynthetically Active Radiation (APAR)," *Proceedings, 6th Symposium on Physical Measurements and Signatures in Remote Sensing,* January 17–21, Val D'Isere, France, 299–306.

Laba, M., Blair, B., Downs, R., Monger, B., Philpot, W., Smith, S., Sullivan, P., and P. Baveye, 2010, "Use of Textural Measurement to Map Invasive Wetland Plants in the Hudson River National Estuarine Research Reserve with INONOS Satellite Imagery," *Remote Sensing of Environment*, 114:876–886.

Lark, R. J., 1996, "Geostatistical Description of Texture on an Aerial Photograph for Discriminating Classes of Land Cover," *International Journal of Remote Sensing*, 17(11):2115–2133.

Li, Q., Cao, X., Jia, K., Zhang, M., and Q. Dong, 2014, "Crop Type Identification by Integration of High-Spatial Resolution Multispectral Data with Features Extracted from Coarse-Resolution Time-Series Vegetation Index Data," *International Journal of Remote Sensing*, DOI: 10.1080/01431161.2014.943325.

Liu, J., Pattey, E., and G. Jego, 2012, "Assessment of Vegetation Indices for Regional Crop Green LAI Estimation from Landsat Images Over Multiple Growing Seasons," *Remote Sensing of Environment*, 123:347–358.

Lo, C. P., and A. K. Yeung, 2007, *Concepts and Techniques of Geographic Information Systems*, 2nd Ed., Upper Saddle River: Prentice-Hall, 492 p.

Luo, L., and G. Mountrakis, 2010, "Integrating Intermediate Inputs from Partially Classified Images within a Hybrid Classification Framework: An Impervious Surface Estimation Example," *Remote Sensing of Environment*, 114:1220–1229.

Lutes, D. C., Keane, R. E., Caratti, J. F., Key, C. H., Benson, N. C., Sutherland, S., and L. J. Gangi, 2006, *FIREMON: Fire Effects Monitoring and Inventory System,* (CD), Fort Collins, CO: USDA, Forest Service, Rocky Mountain Research Station.

Maillard, P., 2003, "Comparing Texture Analysis Methods through Classification," *Photogrammetric Engineering & Remote Sensing*, 69(4):357–367.

Mand, P., and 15 co-authors, 2010, "Responses of the Reflectance Indices PRI and NDVI to Experimental Warming and Drought in European Shrub Lands along a North-South Climatic Gradient," *Remote Sensing of Environment*, 114:626–636.

Mberego, S., Sanga-Ngoie, K., and S. Kobayashi, 2013, "Vegetation Dynamics of Zimbabwe Investigated using NOAA-AVHRR NDVI from 1982 to 2006: A Principal Component Analysis," *International Journal of Remote Sensing*, 34(19):6764–6779.

McGarigal, K., Cushman, S. A., and E. Ene, 2013, *FRAGSTATS, Version 4.0: Spatial Pattern Analysis Program for Categorical and Continuous Maps.* Computer software program produced at the University of Massachusetts, Amherst. Available from http://www.umass.edu/landeco /research/fragstats/fragstats.html.

Mitternicht, G. I., and J. A. Zinck, 2003, "Remote Sensing of Soil Salinity: Potentials and Constraints," *Remote Sensing of Environment*, 85:1–20.

Mumby, P. J., and A. J. Edwards, 2002, "Mapping Marine Environments with IKONOS Imagery: Enhanced Spatial Resolution Can Deliver Greater Thematic Accuracy," *Remote Sensing of Environment*, 82:248–257.

Myers, V. I., 1970, "Soil, Water and Plant Relations," *Remote Sensing with Special Reference to Agriculture and Forestry*, Washington: National Academy of Sciences, 253–297.

Neigh, C. S., Tucker, C. J., and J. R. G. Townshend, 2008, "North American Vegetation Dynamics Observed with Multi-resolution Satellite Data," *Remote Sensing of Environment*, 112:1749–1772.

Nieminen, A., Heinonen, P., and Y. Nuevo, 1987, "A New Class of Detail Preserving Filters for Image Processing," *IEEE Transactions in Pattern Analysis & Machine Intelligence*, 9:74–90.

O'Neill, R. V., Hunsaker, C. T., Jones, K. B., Ritters, K. H., Wickham, J. D, Schwarz, P., Goodman, I. A., Jackson, B., and W. Baillargeon, 1997, "Monitoring Environmental Quality at the Landscape Scale," *BioScience,* 47(8):513–519.

Pena-Barragan, J. M., Ngugi, M. K., Plant, R. E., and J. Six, 2011, "Object-based Crop Classification using Multiple Vegetation Indices, Textural features and Crop Phenology," *Remote Sensing of Environmen*t, 115:1301–1316.

Perry, C. R., and L. F. Lautenschlager, 1984, "Functional Equivalence of Spectral Vegetation Indices," *Remote Sensing of Environment*, 14:169–182.

Peterson, D. L., and S. W. Running, 1989, "Applications in Forest Science and Management," in *Theory and Applications of Optical Remote Sensing*, New York: John Wiley & Sons, 4210–4273.

Pflugmacher, D., Cohen, W. B., and R. E. Kennedy, 2012, "Using Landsat-derived Disturbance History (1972–2010) to Predict Current Forest Structure," *Remote Sensing of Environment*, 122:146–165.

Pierce, L. L., Running, S. W., and G. A. Riggs, 1990, "Remote Detection of Canopy Water Stress in Coniferous Forests Using NS001 Thematic Mapper Simulator and the Thermal Infrared Multispectral Scanner," *Photogrammetric Engineering & Remote Sensing*, 56(5):571–586.

Powell, S. L., Cohen, W. B., Healey, S. P., Kennedy, R. E., Moisen, G. G., Pierce, K. B., et al., 2010, "Quantification of Live Aboveground Forest Biomass Dynamics with Landsat Time-series and Field Inventory Data: A Comparison of Empirical Modeling Approaches," *Remote Sensing of Environment*, 114:1053–1068.

Pratt, W. K., 2014, *Digital Image Processing*, 4th Ed., Boca Raton: CRC Press, 756 p.

Price, K. P., Guo, X., and J. M. Stiles, 2002, "Optimal Landsat TM Band Combinations and Vegetation Indices for Discrimination of Six Grassland Types in Eastern Kansas," *International Journal of Remote Sensing*, 23:5031–5042.

Purkis, S., and V. Klemas, 2011, *Remote Sensing and Global Environmental Change*, New York: Wiley-Blackwell, 367 p.

Purkis, S. J., Myint, S. W., and B. M. Riegl, 2006, "Enhanced Detection of the Coral *Acropora cervcornis* from Satellite Imagery using a Textural Operator," *Remote Sensing of Environment*, 101:82–94.

Qi, J., Cabot, F., Moran, M. S., and G. Dedieu, 1995, "Biophysical Parameter Estimations Using Multidirectional

Spectral Measurements," *Remote Sensing of Environment*, 54:71–83.

Ramsey, R. D., Falconer, A., and J. R. Jensen, 1995, "The Relationship between NOAA-AVHRR NDVI and Ecoregions in Utah," *Remote Sensing of Environment*, 3:188–198.

Richards, J. A., 2013, *Remote Sensing Digital Image Analysis,* 5th Ed., New York: Springer-Verlag, 494 p.

Richardson, A. J., and C. L. Wiegand, 1977, "Distinguishing Vegetation from Soil Background Information," *Remote Sensing of Environment*, 8:307–312.

Ritters, K. H., O'Neill, R. V., Hunsaker, C. T., Wickham, J. D., Yankee, D. H., Timmins, S. P., Jones, K. B., and B. L. Jackson, 1995, "A Factor Analysis of Landscape Pattern and Structure Metrics," *Landscape Ecology*, 10(1):23–39.

Rodriguez-Galiano, V., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., and C. Jeganathan, 2012, "Random Forest Classification of Mediterranean Land Cover using Multi-seasonal Imagery and Multi-seasonal Texture," *Remote Sensing of Environment*, 121:93–107.

Rogan, J., Franklin, J., and D. A. Roberts, 2002, "A Comparison of Methods for Monitoring Multitemporal Vegetation Change Using Thematic Mapper Imagery," *Remote Sensing of Environment*, 80:143–156.

Rondeaux, G., Steven, M., and F. Baret, 1996, "Optimization of Soil-adjusted Vegetation Indices," *Remote Sensing of Environment*, 55:95–107.

Rouse, J. W., Haas, R. H., Schell, J. A., and D. W. Deering, 1974, "Monitoring Vegetation Systems in the Great Plains with ERTS, *Proceedings, 3rd Earth Resource Technology Satellite (ERTS) Symposium,* Vol. 1, 48–62.

Rundquist, B. C, 2002, "The Influence of Canopy Green Vegetation Fraction on Spectral Measurements over Native Tallgrass Prairie," *Remote Sensing of Environment*, 81:129–135.

Running, S. W., Justice, C. O., Solomonson, V., Hall, D., Barker, J., Kaufmann, Y. J., Strahler, A. H., Huete, A. R., Muller, J. P., Vanderbilt, V., Wan, Z. M., Teillet, P., and D. Carneggie, 1994, "Terrestrial Remote Sensing Science and Algorithms Planned for EOS/MODIS," *International Journal of Remote Sensing*, 15(17):3587–3620.

Russ, J. C., 2011, *The Image Processing Handbook*, 6th Ed., Boca Raton: CRC Press, 849 p.

Schlerf, M., Atzberger, C., and J. Hill, 2005, "Remote Sensing of Forest Biophysical Variables using HyMap Imaging Spectrometer Data," *Remote Sensing of Environment*, 95:177–194.

Schowengerdt, R. A., 2007, *Remote Sensing: Models and Methods for Image Processing*, 3rd Ed., San Diego, CA: Academic Press, 515 p.

Sheffield, C., 1985, "Selecting Band Combinations from Multispectral Data," *Photogrammetric Engineering & Remote Sensing*, 51(6):681–687.

Short, N., 1982, "Principles of Computer Processing of Landsat Data," Appendix A in *Landsat Tutorial Workbook*, Publication #1078, Washington: NASA, 421–453.

Sims, D. A., Rahman, A. F., Vermote, E. F., and Z. Jiang, 2011, "Season and Inter-annual Variation in View Angle Effects on MODIS Vegetation Indices at Three Forest Sites," *Remote Sensing of Environment*, 115:3112–3120.

Small, C., 2012, "Spatiotemporal Dimensionality and Time-Space Characterization of Multitemporal Imagery," *Remote Sensing of Environment*, 124:793–809.

Sonnenschein, R., Kuemmerle, T., Udelhoven, T., Stellmes, M., and P. Hostert, 2011, "Differences in Landsat-based Trend Analyses in Drylands due to the Choice of Vegetation Estimate," *Remote Sensing of Environment*, 115:1408–1420.

TBRS, 2003, *Enhanced Vegetation Index,* Terrestrial Biophysics and Remote Sensing Lab, University of Arizona, http://tbrs.arizona.edu/project/MODIS/evi.php.

Troll, C., 1939, Luftbildplan and okologische Bodenforschung, *A. Ges. Erdkunde*, Berlin: 241–298.

Tucker, C. J., 1979, "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation," *Remote Sensing of Environment*, 8:127–150.

Tuominen, S., and A. Pekkarinen, 2005, "Performance of Different Spectral and Textural Aerial Photograph Features in Multi-source Forest Inventory," *Remote Sensing of Environment*, 94:256–268.

Verbesselt, J., Somers, B., Lhermitte, S., Jonckheere, I., van Aardt, J., and P. Coppin, 2007, "Monitoring Herbaceous Fuel Moisture Content with SPOT VEGETATION Time-series for Fire Risk Prediction in Savannah Ecosystems," *Remote Sensing of Environment*, 108:357–368.

Vina, A., Gitelson, A. A., Nguy-Robertson, A. L., and Y. Peng, 2011, "Comparison of Different Vegetation Indices for the Remote Assessment of Green Leaf Area Index of Crops," *Remote Sensing of Environment*, 115:3468–3478.

Viscarra-Rossel, R. A., and C. Chen, 2011, "Digitally Mapping the Information Content of Visible-Near Infrared Spectra of Surficial Australian Soils," *Remote Sensing of Environment*, 115:1443–1455.

Wagle, P., and 8 co-authors, 2014, "Sensitivity of Vegetation Indices and Gross Primary Production of Tallgrass Prairie to Severe Drought," *Remote Sensing of Environment*, 152:1–14.

Walter-Shea, E. A., and L. L. Biehl, 1990, "Measuring Vegetation Spectral Properties," *Remote Sensing Reviews*, 5(1):179–205.

Wang, F., 1993, "A Knowledge-based Vision System for Detecting Land Changes at Urban Fringes," *IEEE Transactions on Geoscience and Remote Sensing*, 31(1):136–145.

Wang, L. and D. C. He, 1990, "A New Statistical Approach for Texture Analysis," *Photogrammetric Engineering & Remote Sensing*, 56(1):61–66.

Wang, L., Hunt, E. R., Qu, J. J., Hao, X., and C. W. Daughtry, 2013, "Remote Sensing of Fuel Moisture Content from Ratios of Narrow-band Vegetation Water and Dry-matter Indices," *Remote Sensing of Environment*, 129:103–110.

Wang, L., Sousa, W., Gong, P., and G. Biging, 2004, "Comparison of IKONOS and QuickBird Images for Mapping Mangrove Species on the Caribbean Coast of Panama," *Remote Sensing of Environment*, 91:432–440.

Wang, L., and S. Zhang, 2014, "Incorporation of Texture Information in a SVM Method for Classifying Salt Cedar in Western China," *Remote Sensing Letters*, 5(9):501–510.

Wardlow, B. D., Egbert, S. L., and J. H. Kastens, 2007, "Analysis of Time-series MODIS 250 m Vegetation Index Data for Crop Classification in the U.S. Central Great Plains," *Remote Sensing of Environment*, 108:290–310.

Warner, T., 2011, "Kernel-Based Texture in Remote Sensing Image Classification," *Geography Compass*, 5/10:781–798.

Warner, T. A., Almutairi, A., and J. Y. Lee, 2009, "Remote Sensing of Land Cover Change," in Warner, T. A., Nellis, M. D., and G. M. Foody (Eds.), *The Sage Handbook of Remote Sensing*, Los Angeles, Sage, Inc., 459–472.

Weiers, S., Bock, M., Wissen, M., and G. Rossner, 2004, "Mapping and Indicator Approaches for the Assessment of Habitats at Different Scales using Remote Sensing and GIS Methods," *Landscape and Urban Planning*, 67:43–65.

Wickham, J. D., and D. J. Norton, 1994, "Mapping and Analyzing Landscape Patterns," *Landscape Ecology*, 9(1):7–23.

Wood, E. M., Pidgeon, A. M., Radeloff, V. C., and N. S. Keuler, 2012, "Image Texture as a Remotely Sensed Measure of Vegetation Structure," *Remote Sensing of Environment*, 121:516–526.

Woodcock, C. E., Strahler, A. H., and D. L. B. Jupp, 1988, "The Use of Variogram in Remote Sensing and Simulated Image, II: Real Digital Images," *Remote Sensing of Environment*, 25:349–379.

Yarbrough, L. D., Easson, G., and J. S. Juszmaul, 2012, "Proposed Workflow for Improved Kauth-Thomas Transform Derivations," *Remote Sensing of Environment*, 124:810–818.

Yarbrough, L. D., Navulur, K., and R. Ravi, 2014, "Presentation of the Kauth-Thomas Transform for WorldView-2 Reflectance Data," *Remote Sensing Letters*, 5(2):131–138.

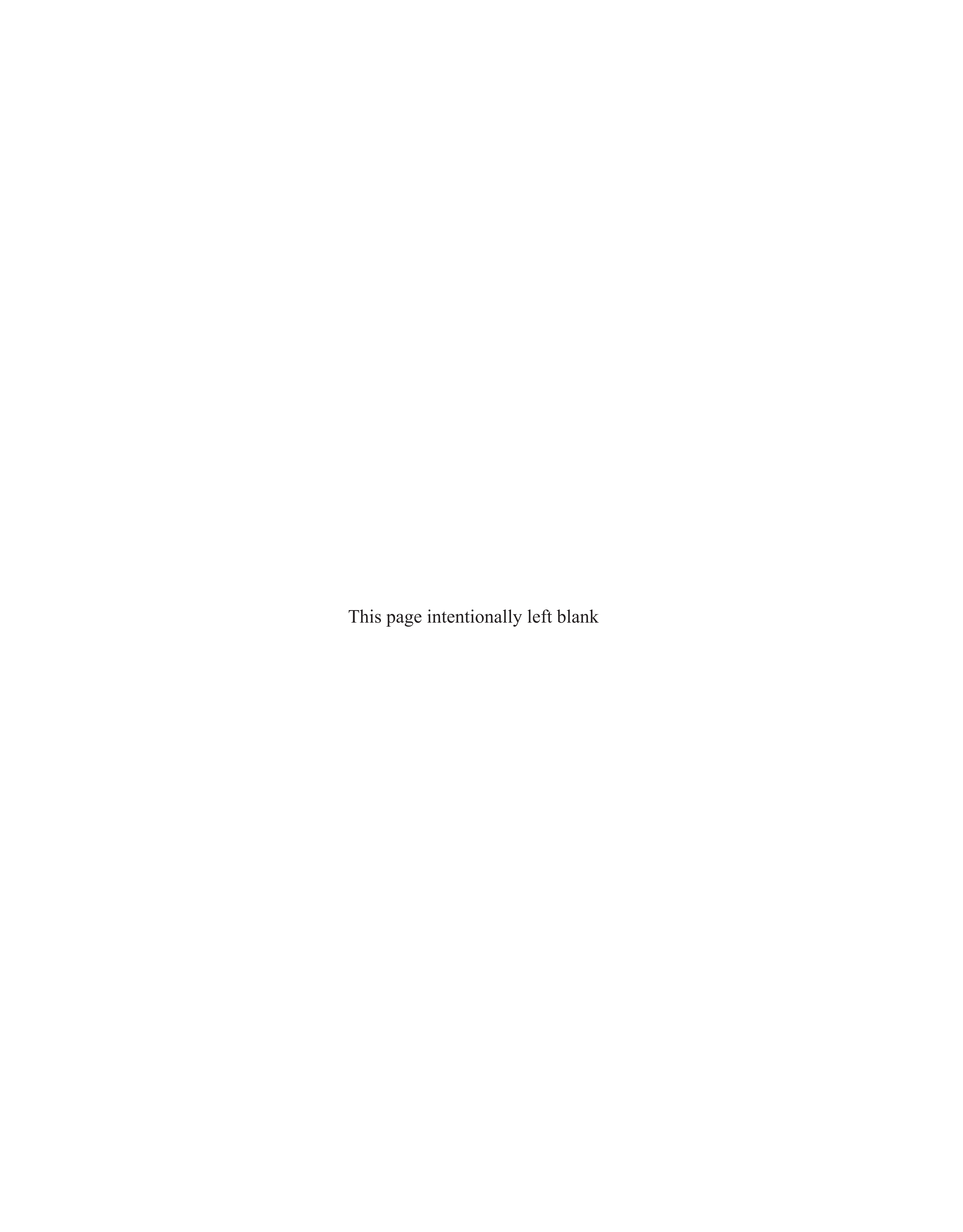
Zha, Y., Gao, J., and S. Ni, 2003, "Use of Normalized Difference Built-up Index in Automatically Mapping Urban Areas from TM Imagery," *International Journal of Remote Sensing*, 24(3):583–594.

Zhang, C., and Z. Xie, 2012, "Combining Object-based Texture Measures with a Neural Network for Vegetation Mapping in the Everglades from Hyperspectral Imagery," *Remote Sensing of Environment*, 124:310–320.

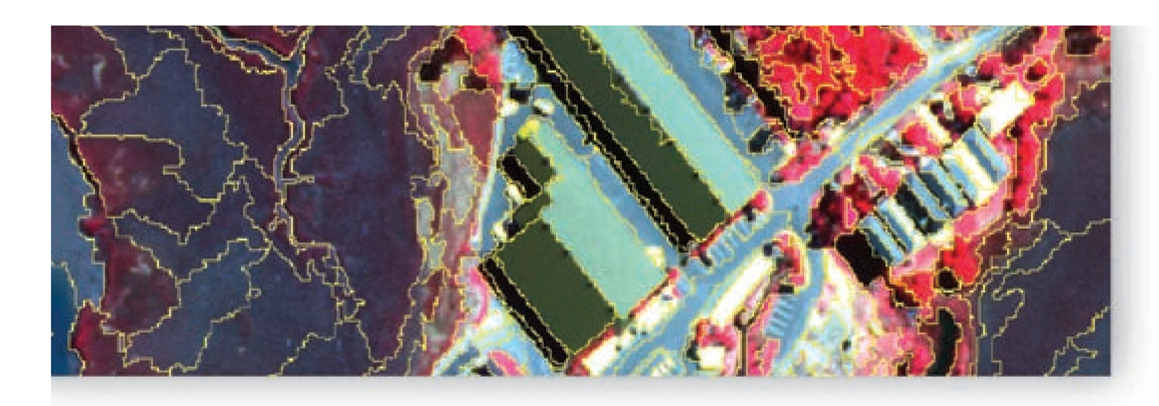
Zhang, Y., 2001, "Texture-Integrated Classification of Urban Treed Areas in High-resolution Color-infrared Imagery," *Photogrammetric Engineering & Remote Sensing*, 67(12):1359–1365.

Zhang, Z., Schaaf, C., and K. C. Seto, 2013, "The Vegetation Adjusted NTL Urban Index: A New Approach to Reduce Saturation and Increase Variation in Nighttime Luminosity," *Remote Sensing of Environment*, 129:32–41.

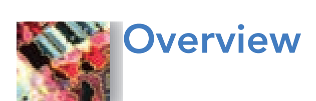
Zhao, G., and A. L Maclean, 2000, "A Comparison of Canonical Discriminant Analysis and Principal Component Analysis for Spectral Transformation," *Photogrammetric Engineering & Remote Sensing*, 66(7):841–847.


This page intentionally left blank

9 THEMATIC INFORMATION EXTRACTION: PATTERN RECOGNITION

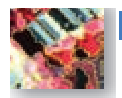


Remote sensing data are collected by an increasingly diverse array of airborne and satellite remote sensing systems (e.g., Colomina and Molina, 2014; Belward and Skoien, 2014). The *data* can be processed and transformed into various types of thematic *information*. In particular, remote sensing has increasingly been used as a source of land-use and land-cover information at local, regional, and global scales (Homer et al., 2012; Jensen and Jensen, 2013). The extraction of thematic information is often performed by identifying spectral and/or spatial patterns in the remote sensor data that are correlated with different types of land-use or land-cover present. The term **pattern recognition** is commonly used to describe this analytical process.



Overview

This chapter reviews the fundamental methods used to apply pattern recognition techniques to multispectral data. It begins by reviewing the general steps required to extract land-cover information from digital remote sensor data. The concept of *hard* (crisp) versus *soft* (fuzzy) classification logic is introduced. The characteristics of several of the most important land-cover classification schemes are reviewed. Numerous methods of supervised classification (e.g., parallelepiped, minimum distance, maximum likelihood) and unsupervised clustering (e.g., ISODATA) are discussed. Feature selection methods used to identify the most appropriate bands for use in the classification algorithms are described. The chapter concludes with a discussion of object-based image analysis (OBIA) procedures that are used to extract homogeneous patches (polygons) of information from the remote sensor data instead of thematic information in pixel format. Nonmetric methods including neural networks and expert system decision-tree classifiers are introduced in Chapter 10. The algorithms required to extract information from hyperspectral imagery are discussed in Chapter 11.



Introduction

Pattern recognition is normally applied to multispectral remote sensor data that has been geometrically corrected. General steps required to extract land-cover information from digital multispectral remote sensor data are summarized in Figure 9-1.

Multispectral classification may be performed using a variety of methods, including (Figures 9-1 and 9-2):

- algorithms based on *parametric* and *nonparametric* statistics that use ratio- and interval-scaled data and *nonmetric* methods that can also incorporate nominal scale data,
- the use of *supervised* or *unsupervised* classification logic,
- use of *hard* (crisp) or *soft* (fuzzy) classification logic,
- use of per-pixel or object-based classification logic, and
- *hybrid* approaches.

Parametric methods such as maximum likelihood classification and unsupervised clustering assume normally distributed remote sensor data and knowledge about the forms of the underlying class density functions (Duda et al., 2001). *Nonparametric* methods such as

nearest-neighbor classifiers, fuzzy classifiers, and neural networks may be applied to remote sensor data that are not normally distributed and without the assumption that the forms of the underlying densities are known. *Nonmetric* methods such as rule-based decision-tree classifiers can operate on both real-valued data (e.g., reflectance values from 0 to 100%) and nominal scaled data (e.g., class 1 = forest; class 2 = agriculture) (e.g., Jensen et al., 2009; Myint et al., 2011).

In a **supervised classification**, the identity and location of some of the land-cover types (e.g., urban, agriculture, or wetland) are known *a priori* (i.e., before) through a combination of fieldwork, interpretation of aerial photography, map analysis, and personal experience. The analyst attempts to locate specific sites in the remotely sensed data that represent homogeneous examples of these known land-cover types. These areas are commonly referred to as **training sites** because the spectral characteristics of these known areas are used to train the classification algorithm. Multivariate statistical parameters (means, standard deviations, covariance matrices, correlation matrices, etc.) are calculated for each training site. Every pixel both within and outside the training sites is then evaluated and assigned to the class of which it has the highest likelihood of being a member.

In an **unsupervised classification**, the identities of land-cover types to be specified as classes within a scene are not generally known *a priori* because ground reference information is lacking or surface features within the scene are not well defined. The computer is required to group pixels with similar spectral characteristics into unique clusters according to some statistically determined criteria. The analyst then relabels and combines the spectral clusters into information classes.

Supervised and unsupervised classification algorithms typically use *hard* (crisp) *classification* logic to produce a classification map that consists of hard, discrete categories (e.g., forest, agriculture) (Figure 9-2a). Conversely, it is also possible to use *fuzzy* (soft) *classification* logic, which takes into account the heterogeneous and imprecise nature of the real world (Figure 9-2b). Fuzzy classification produces thematic output products that contain fuzzy information. Fuzzy classification is based on the fact that remote sensing detectors record the reflected or emitted radiant flux from heterogeneous mixtures of biophysical materials such as soil, water, and vegetation found within the IFOV. The land-cover classes found within the IFOV (pixel) often grade into one another without sharp, hard boundaries. Thus, reality is actually very imprecise and heterogeneous; that is, it is fuzzy (Jensen et al., 2009). Instead of being assigned to just a single class out of *m* possible classes, each pixel in a fuzzy classification has *m* membership grade values that describe the proportion of the *m* land-cover types found within the pixel (e.g., 10% bare soil, 10% scrub-shrub, 80% forest) (Figure 9-2b). This information may be used to extract more precise land-cover information, especially concerning the makeup of mixed pixels (Foody, 2002).

In the past, most digital image classification was based on processing the entire scene pixel by pixel. This is commonly referred to as **per-pixel classification** (Blaschke and Strobl, 2001; Myint et al., 2011). **Geographic object-based image analysis** (GEOBIA) techniques allow the analyst to decompose the scene into many relatively homogenous image objects (referred to as patches or segments) using a multiresolution image segmentation process (Blaschke et al., 2014). The various statistical characteristics of these homogeneous image objects in the scene are then subjected to traditional statistical or fuzzy logic classification. Object-based classification based on image segmentation is often used for the analysis of high spatial resolution imagery (e.g., GeoEye-1, Pleiades, WorldView-2).

No pattern classification method is inherently superior to any other. The nature of the classification problem, the biophysical characteristics of the study area, the distribution of the remotely sensed data (e.g., normally distributed), and *a priori* knowledge determine which classification algorithm will yield useful results (Duro et al., 2012). Duda et al. (2001) provide sound advice: "We should have a healthy skepticism regarding studies that purport to demonstrate the overall superiority of a particular learning or recognition algorithm."



Supervised Classification

Useful thematic information may be obtained using supervised classification algorithms if the general steps summarized in Figure 9-1 are understood and applied. The analyst first specifies the geographic **region of interest** (ROI) on which to test hypotheses. The classes of interest to be examined are then carefully defined in a classification scheme. The classes of interest normally dictate whether the analyst should produce hard or fuzzy output products and whether per-pixel or object-oriented classification logic should be used. Next, the analyst obtains the appropriate digital remote sensor data, keeping in mind both sensor system and environmental constraints. Ideally ground reference information is obtained at the same as the remote sensing data acquisition. The remote sensor data are radiometrically and geometrically corrected as discussed in Chapters 6 and 7. An appropriate classification algorithm is then selected and the analyst selects initial training data (if necessary). Feature (band) selection is performed to

General Steps Used to Extract Thematic Land-Cover Information from Digital Remote Sensor Data

State the nature of the land-cover classification problem.
 * Specify the geographic region of interest (ROI).
 * Define the classes of interest within a classification scheme.
 * Determine if it is to be a *hard* (crisp) or *soft* (fuzzy) classification.
 * Determine if it is to be a per-pixel or object-based (OBIA) classification.
Acquire appropriate remote sensing and initial ground reference data.
 * Select remotely sensed data based on the following criteria:
  - Remote sensing system considerations:
   - Spatial, spectral, temporal, and radiometric resolution
  - Environmental considerations:
   - Atmospheric, soil moisture, phenological cycle, etc.
 * Obtain initial ground reference data based on:
  - *a priori* knowledge of the study area
Process remote sensor data to extract thematic information.
 * Radiometric correction (or normalization) if necessary (Chapter 6)
 * Geometric correction (Chapter 7)
 * Select appropriate image classification logic:
  - Parametric (e.g., maximum likelihood, clustering)
  - Nonparametric (e.g., nearest-neighbor, neural network)
  - Nonmetric (e.g., rule-based decision-tree classifier)
 * Select appropriate image classification algorithm:
  - Supervised, e.g.,
   - Parallelepiped, minimum distance, maximum likelihood
   - Others (hyperspectral matched filtering, spectral angle mapper – Chapter 11)
  - Unsupervised, e.g.,
   - Chain method, multiple-pass ISODATA
   - Others (fuzzy *c*-means)
  - Hybrid involving artificial intelligence (Chapter 10), e.g.,
   - Expert system decision-tree, neural network, support vector machine
 * Extract data from initial training sites (if required).
 * Select the most appropriate bands using feature selection criteria:
  - Graphical (e.g., two-dimensional feature space plots)
  - Statistical (e.g., transformed divergence, TM-distance)
 * Extract training statistics and rules based on:
  - Final band selection (if required), and/or
  - Machine-learning (Chapter 10)
 * Extract thematic information:
  - For each pixel or for each OBIA image object (supervised)
  - Label pixels or image objects (unsupervised)
Accuracy assessment (Chapter 13).
 * Select method:
  - Qualitative confidence-building
  - Statistical measurement
 * Determine number of samples required by class.
 * Select sampling scheme.
 * Obtain ground reference test information.
 * Create and analyze error matrix:
  - Univariate and multivariate statistical analysis.
Accept or reject previously stated hypothesis.
Distribute results if accuracy is acceptable.

FIGURE 9–1 The general steps used to extract thematic land-cover information from digital remote sensor data.

Classification of Remotely Sensed Data Based on Hard (Crisp) versus Soft (Fuzzy) Logic

Single-Stage Hard Classification of One Pixel to One Class

1
2
3
n
Rectified remotely sensed data

Hard partition of feature space and assignment of each pixel to one of *m* classes using supervised and/or unsupervised classification logic.

Final classification map consisting of *m* classes.

a.

Computation of Fuzzy Membership Grades and Final Classification

1
2
3
n
Rectified remotely sensed data

Fuzzy partition of feature space where each pixel has a membership grade value (from 0 to 1) for *m* classes using supervised and/or unsupervised classification logic.

1
2
3
4
5
6
m

Application of additional logic to the membership grade information to derive a final classification map consisting of *m* classes, if desired.

b.

FIGURE 9–2 Relationship between a traditional single-stage hard (crisp) classification using supervised or unsupervised classification logic and a classification based on the use of soft (fuzzy) logic.

determine the multispectral bands optimal for discriminating one training class from another. Additional training data are collected (if necessary) and the classification algorithm is applied, yielding a classification map. A rigorous accuracy assessment (often referred to as an error evaluation) is then performed (see Chapter 13). The classification maps and associated statistics are then distributed to colleagues and agencies when the results are acceptable.

Land-Use and Land-Cover Classification Schemes

Land cover refers to the type of material present on the landscape (e.g., water, crops, forest, wetland, human-made materials such as asphalt). **Land use** refers to what people do on the land surface (e.g., agriculture, commerce, settlement). The pace, magnitude, and scale of human alterations of the Earth's land surface are unprecedented in human history. Consequently, land-cover and land-use data are central to such United Nations' *Agenda 21* issues as combating deforestation, managing sustainable settlement growth, and protecting the quality and supply of water resources (Jensen et al., 2002). In light of the human impacts on the landscape, there is a need to establish baseline datasets against which changes in land cover and land use can be assessed (Warner et al., 2009; Weng, 2014).

The International Geosphere-Biosphere Programme (IGBP) and the International Human Dimensions of Global Environmental Change Programme (IHDP) suggest that:

> over the coming decades, the global effects of land use and cover change may be as significant, or more so, than those associated with potential climate change. Unlike climate change per se, land use and cover change are known and undisputed aspects of global environmental change. These changes and their impacts are with us now, ranging from potential climate warming to land degradation and biodiversity loss and from food production to spread of infectious diseases (IGBP-IHDP, 2002).

Land-cover data have proved especially valuable for predicting the distribution of both individual species and species assemblages across broad areas that could not otherwise be surveyed. Various predictive models have gained currency as the availability and accuracy of land-cover datasets have improved. For example, remote sensing–derived land-cover information is used extensively in the Gap Analysis Program (GAP), which is the largest species distribution modeling effort. The goal is to develop detailed maps of habitat preferences for target species and monitor plant phenology (Kerr and Ostrovsky, 2003).

All classes of interest must be selected and defined carefully to classify remotely sensed data successfully into land-use and/or land-cover information (Congalton and Green, 2009). This requires the use of a **classification scheme** containing taxonomically correct definitions of classes of information that are organized according to logical criteria. If a hard (crisp) classification is to be performed, then the classes in the classification system should normally be:

- mutually exclusive,
- exhaustive, and
- hierarchical.

Mutually exclusive means that there is no taxonomic overlap (or fuzziness) of any classes (i.e., deciduous forest and evergreen forest are distinct classes). *Exhaustive* means that all land-cover classes present in the landscape are accounted for and none has been omitted. *Hierarchical* means that sublevel classes (e.g., single-family residential, multiple-family residential) may be hierarchically combined into a higher-level category (e.g., residential) that makes sense. This allows simplified thematic maps to be produced when required.

It is also important for the analyst to realize that there is a fundamental difference between information classes and spectral classes. ***Information classes*** are those that human beings define. Conversely, ***spectral classes*** are those that are inherent in the remote sensor data and must be identified and then labeled by the analyst. For example, in a remotely sensed image of an urban area there is likely to be single-family residential housing. A relatively coarse spatial resolution remote sensor such as SPOT (20×20 m) might be able to record a few relatively pure pixels of vegetation and a few pure pixels of concrete/asphalt road or asphalt shingles. However, it is more likely that in this residential area the pixel brightness values will be a function of the reflectance from mixtures of vegetation, asphalt and concrete. Few planners or administrators want to see a map labeled with classes like (1) concrete, (2) asphalt, (3) vegetation, and (4) mixture of vegetation and concrete/asphalt (unless they are interested in a map of impervious surfaces). Rather, they typically prefer the analyst to relabel the mixture class as single-family residential. The analyst should do this only if in fact there is a good association between the mixture class and single-family residential housing. Thus, we see that an analyst must often translate *spectral classes* into *information classes* to satisfy bureaucratic requirements. An analyst should understand well the spatial and spectral characteristics of the sensor system and be able to relate these system parameters to the types and proportions of materials found within the scene and within pixel IFOVs. If these parameters and relationships are understood, spectral classes often can be thoughtfully relabeled as information classes.

Several hard classification schemes have been developed that can readily incorporate land-use and/or land-cover data obtained by interpreting remotely sensed data. Only a few of these classification schemes will be discussed here, including:

- American Planning Association (APA) *Land-Based Classification Standard,* which is oriented toward detailed land-use classification (Figure 9-3);
- U.S. Geological Survey *Land-Use/Land-Cover Classification System for Use with Remote Sensor Data;*
- U.S. National Land Cover Dataset (NLCD) Classification scheme (Figure 9-4);
- NOAA Coastal Change Analysis Program (C-CAP) classification scheme;
- U.S. Department of the Interior Fish & Wildlife Service*, Classification of Wetlands and Deepwater Habitats of the United States* (Figures 9-5 and 9-6);
- U.S. *National Vegetation & Classification Standard*;
- International Geosphere-Biosphere Program *IGBP Land Cover Classification System* specially modified for the creation of MODIS land-cover products (Figure 9-7).

American Planning Association *Land-Based Classification Standard (LBCS)*

Few classification schemes attempt to classify land use. In fact, most explicitly state that they are concerned only with land-cover information. Therefore, if the user is most interested in extracting detailed urban/suburban land-use information from relatively high spatial resolution remote sensor data, then one of the most practical and comprehensive hierarchical classification systems is the *Land-Based Classification Standard* (LBCS) developed by the American Planning Association (2014a).

The LBCS requires input from *in situ* surveys, and/or aerial photography, and satellite remote sensor data to obtain information at the parcel level on the following five characteristics: *activity, function, site development, structure,* and *ownership* (American Planning Association, 2014a). The system provides a unique code and description for almost every commercial and industrial land-use activity (APA, 2014b). An example of the *Land-Based Classification Standard* applied to three parcels of land in Beaufort, SC, in 2011, is shown in Figure 9-3 along with their associated LBCS activity, function, and structure codes. Two of the commercial establishments are full-service fast-food restaurants with drive-throughs and the other is a furniture store.

American Planning Association *Land-Based Classification Standard* (LBCS)



Table 1: Activity

Parcel ID	Activity	Description
0171	2210	restaurant with drive-through
0170	2100	furniture
125A	2210	restaurant with drive-through

Table 2: Function

Function	Description
2510	full-service restaurant
2121	furniture
2510	full-service restaurant

Table 3: Structural Character

Parcel ID	Structure	Description
0171	2220	restaurant building
0170	2592	home improvement center
125A	2220	restaurant building

FIGURE 9–3 The American Planning Association developed the *Land-Based Classification Standard* (LBCS) that contains detailed definitions of urban/suburban land use. The system incorporates information derived *in situ* and using remote sensing techniques. This is an oblique aerial photograph of three commercial buildings in Beaufort, SC, in 2011. Activity, function, and structural character codes associated with these three parcels are identified. Site development and ownership information attribute tables are not shown (aerial photography and parcel information courtesy of Beaufort County GIS Department).

The LBCS is always under development. Users are encouraged to keep abreast of the LBCS and to use it for very intensive urban/suburban studies that require detailed commercial and industrial land-use classification codes. The LBCS does not provide information on land-cover or vegetation characteristics in the urban environment; it relies on the Federal Geographic Data Committee standards on this topic.

U.S. Geological Survey Anderson *Land-Use/ Land-Cover Classification System for Use with Remote Sensor Data*

The U.S. Geological Survey's *Land-Use/Land-Cover Classification System for Use with Remote Sensor Data* (Anderson et al., 1976) is primarily a resource-oriented land-cover classification system in contrast with people or activity land-use classification systems such as the APA's *Land-Based Classification Standard*. The USGS rationale is that "although there is an obvious need for an urban-oriented land-use classification system, there is also a need for a resource-oriented classification system whose primary emphasis would be the remaining 95 percent of the United States land area." The USGS system addresses this need with eight of the nine original Level I categories that treat land area that is not in urban or built-up categories (Table 9-1). The system is designed to be driven primarily by the interpretation of remote sensor data obtained at various scales and resolutions (Table 9-2) and not data collected *in situ*. The

TABLE 9–1 U.S. Geological Survey *Land-Use/Land-Cover Classification System for Use with Remote Sensor Data* (Anderson et al., 1976).

Classification Level
1 Urban or Built-up Land 11 Residential 12 Commercial and Services 13 Industrial 14 Transportation, Communications, and Utilities 15 Industrial and Commercial Complexes 16 Mixed Urban or Built-up 17 Other Urban or Built-up Land
2 Agricultural Land 21 Cropland and Pasture 22 Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultural Areas 23 Confined Feeding Operations 24 Other Agricultural Land
3 Rangeland 31 Herbaceous Rangeland 32 Shrub–Brushland Rangeland 33 Mixed Rangeland
4 Forest Land 41 Deciduous Forest Land 42 Evergreen Forest Land 43 Mixed Forest Land
5 Water 51 Streams and Canals 52 Lakes 53 Reservoirs 54 Bays and Estuaries
6 Wetland 61 Forested Wetland 62 Nonforested Wetland
7 Barren Land 71 Dry Salt Flats 72 Beaches 73 Sandy Areas Other Than Beaches 74 Bare Exposed Rock 75 Strip Mines, Quarries, and Gravel Pits 76 Transitional Areas 77 Mixed Barren Land
8 Tundra 81 Shrub and Brush Tundra 82 Herbaceous Tundra 83 Bare Ground Tundra 84 Wet Tundra 85 Mixed Tundra
9 Perennial Snow or Ice 91 Perennial Snowfields 92 Glaciers

classification system continues to be modified to support a great variety of land-cover mapping activities by the U.S. Geological Survey, the Environmental Protection Agency, the NOAA Coastal Services Center, the Multi-Resolution Land Characteristics consortium, and others.

TABLE 9–2 Four levels of the U.S. Geological Survey *Land-Use/Land-Cover Classification System for Use with Remote Sensor Data* and the type of remotely sensed data typically used to provide the information (Anderson et al., 1976; Jensen and Cowen, 1999).

Classification Level	Typical Data Characteristics
I	Satellite imagery such as NOAA AVHRR (1.1 × 1.1 km), MODIS (250 × 250 m; 500 × 500 m), Landsat MSS (79 × 79 m), Landsat TM and Landsat 8 (30 × 30 m), and SPOT XS (20 × 20 m).
II	Satellite imagery such as SPOT HRV multispectral (10 × 10 m) and Indian IRS 1-C panchromatic (5 × 5 m). High-altitude aerial photography acquired at scales smaller than 1:80,000.
III	Satellite imagery with 1 × 1 m to 2.5 × 2.5 m nominal spatial resolution such as IKONOS. Medium-altitude aerial photography at scales from 1:20,000 to 1:80,000.
IV	Satellite imagery with ≤ 1 × 1 m nominal spatial resolution (e.g., GeoEye-1; WorldView-2). Low-altitude aerial photography at scales from 1:4,000 to 1:20,000 scale.

National Land Cover Database (NLCD) Classification System

The Multi-Resolution Land Characteristics (MRLC) consortium, a group of federal agencies led by the U.S. Geological Survey, originally joined forces in 1992 to use Landsat Thematic Mapper 30 × 30 m imagery to create a "wall-to-wall" land-cover dataset of the nation called the **National Land Cover Database (NLCD)**. The classification scheme used by the NLCD was adapted from the Anderson Land Cover Classification System and is listed in Table 9-3 (MRLC, 2014). The NLCD consists of three major data releases based on a 10-year production cycle:

- a circa 1992 conterminous U.S. land-cover dataset (referred to as NLCD 1992),
- a circa 2001 land-cover database for the U.S. and Puerto Rico (referred to as NLCD 2001), and
- a 1992/2001 Land Cover Change Retrofit (LCCR) product.

The NLCD 2006 represents the first database produced with a 5-year repeat cycle and with additional products specifically designed for land-cover monitoring (e.g., percent impervious surface and percent tree canopy cover). NLCD 2011 is scheduled for release in 2014 (Homer et al., 2012). The NLCD program has changed its emphasis from mapping to monitoring to address the emerging issues of sustainable use that are

TABLE 9–3 National Land Cover Dataset (NLCD) 2006 Classification Scheme (Homer et al., 2012; MRLC, 2014). The list does not include several Alaskan land-cover classes, including: dwarf scrub, sedge/herbaceous, lichens, and moss.

Class/Value	National Land Cover Database (NLCD) 2006 Classification Scheme
Water	*areas of open water or permanent ice/snow cover.*
11	**Open Water** – areas of open water, generally with <25% cover of vegetation or soil.
12	**Perennial Ice/Snow** – areas characterized by a perennial cover of ice and/or snow, generally >25% of total cover.
Developed	*areas with a high percentage (≥30%) of constructed materials (e.g. asphalt, concrete, buildings, etc.).*
21	**Developed, Open Space** – a mixture of constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for <20% of total cover. Includes large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes.
22	**Developed, Low Intensity** – areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20% to 49% of total cover. These areas most commonly include single-family housing units.
23	**Developed, Medium Intensity** – areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50% to 79% of the total cover. These areas most commonly include single-family housing units.
24	**Developed High Intensity** – highly developed areas where people reside or work in high numbers such as apartments, row houses and commercial/industrial. Impervious surfaces account for 80% to 100% of the total cover.
Barren	*characterized by bare rock, gravel, sand, silt, clay, or other earthen material, with little or no "green" vegetation present regardless of its inherent ability to support life. Vegetation, if present, is more widely spaced and scrubby than that in the green vegetated categories; lichen cover may be extensive.*
31	**Barren Land (Rock/Sand/Clay)** – bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits and other earthen material. Vegetation accounts for <15% of total cover.
Forest	*tree cover (natural or semi-natural woody vegetation, >6 m tall); tree canopy accounts for 25% to 100% of cover.*
41	**Deciduous Forest** – areas dominated by trees generally >5 m tall, and >20% of total vegetation cover. >75% of the tree species shed foliage simultaneously in response to seasonal change.
42	**Evergreen Forest** – areas dominated by trees generally >5 m tall, and >20% of total vegetation cover. >75% of the tree species maintain their leaves all year. Canopy is never without green foliage.
43	**Mixed Forest** – areas dominated by trees generally >5 m tall, and >20% of total vegetation cover. Neither deciduous nor evergreen species are >75% of total tree cover.
Shrubland	*characterized by natural or semi-natural woody vegetation with aerial stems, generally <6 m tall, with individuals or clumps not touching to interlocking. Both evergreen and deciduous species of true shrubs, young trees, and trees or shrubs that are small or stunted because of environmental conditions are included.*
51	**Dwarf Scrub** – Alaska only areas dominated by shrubs <20 cm tall with shrub canopy typically >20% of total vegetation. This type is often co-associated with grasses, sedges, herbs, and non-vascular vegetation.
52	**Shrub/Scrub** – areas dominated by shrubs; <5 m tall with shrub canopy typically >20% of total vegetation. Class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions.
Herbaceous	*characterized by natural or semi-natural herbaceous vegetation that accounts for 75% to 100% of cover.*
71	**Grassland/Herbaceous** – areas dominated by gramanoid or herbaceous vegetation, generally >80% of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.
Planted/Cultivated	*characterized by herbaceous vegetation that has been planted or is intensively managed for the production of food, feed, or fiber. Herbaceous vegetation accounts for 75% to 100% of the cover.*
81	**Pasture/Hay** – grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for >20% of total vegetation.
82	**Cultivated Crops** – areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for >20% of total vegetation. This class also includes all land being actively tilled.
Wetlands	*where the soil or substrate is periodically saturated or covered with water as defined by Cowardin et al. (1979).*
90	**Woody Wetlands** – areas where forest or shrubland vegetation accounts for >20% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.
95	**Emergent Herbaceous Wetlands** – Areas where perennial herbaceous vegetation accounts for >80% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

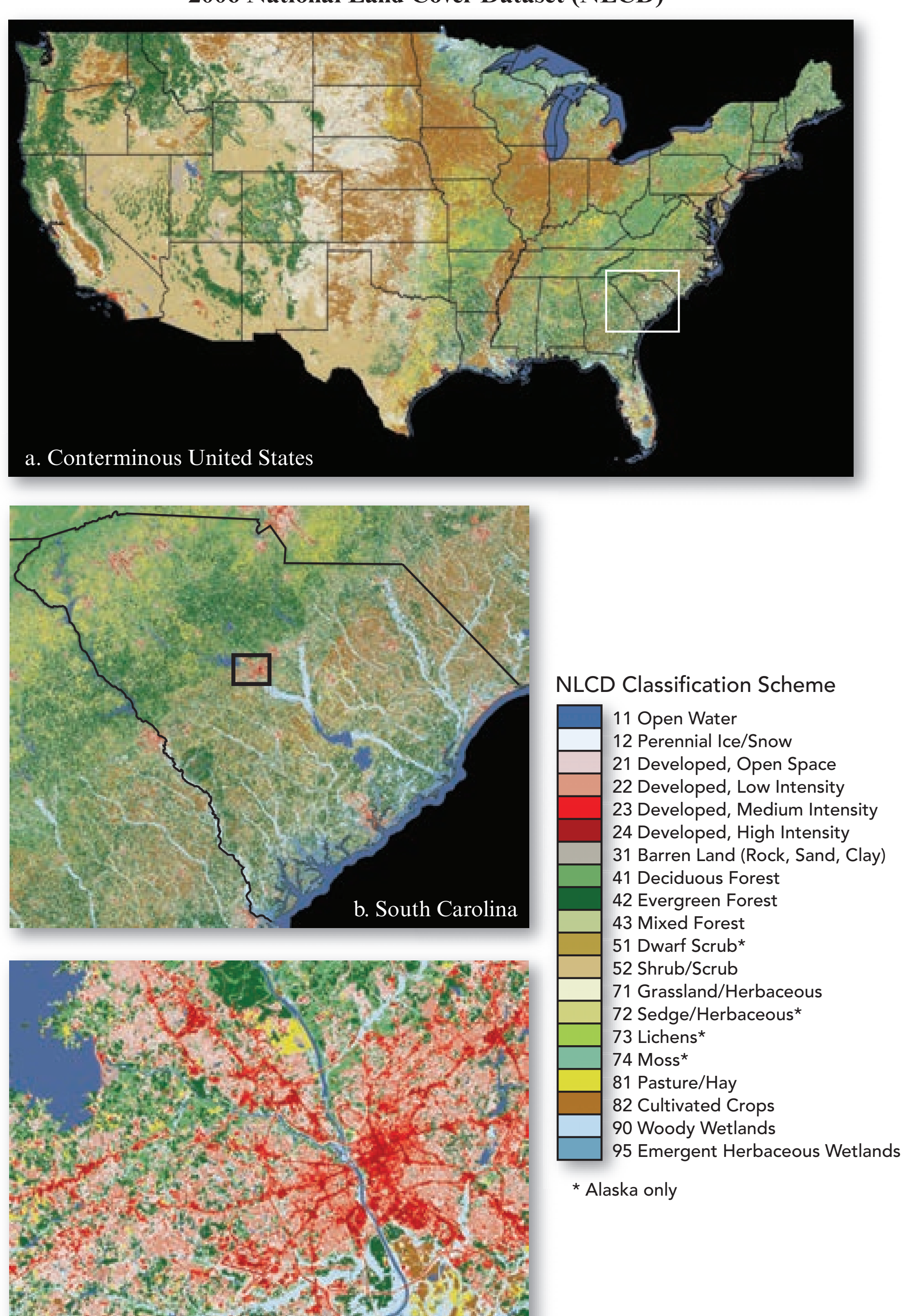


FIGURE 9–4 a) 2006 NLCD land cover information of the conterminous United States. b) 2006 NLCD land-cover information of South Carolina. c) 2006 NLCD land-cover information for Columbia, SC (NLCD information courtesy of the Multi-Resolution Land Characteristics (MRLC) consortium).

faced by the MRLC and its user community (Xian et al., 2009; Fry et al., 2011). NLCD 2006 land-cover information for the conterminous United States, the state of South Carolina, and Columbia, SC, are displayed in Figure 9-4.

TABLE 9–4 NOAA Coastal Change Analysis Program (C-CAP) land cover classification scheme (NOAA C-CAP, 2014).

Class/Value	NOAA C-CAP Land-Cover Classification Description
Developed **Forest** **Shrubland** **Herbaceous** **Planted/Cultivated**	Identical to NLCD classes listed in Table 9-3.
Palustrine Wetlands	
13	**Palustrine Forested Wetland –** includes tidal and nontidal wetlands dominated by woody vegetation ≥5 m in height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is <0.5%. Total vegetation coverage is >20%.
14	**Palustrine Scrub/Shrub Wetland** – includes tidal and nontidal wetlands dominated by woody vegetation <5 m in height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is <0.5%. Total vegetation coverage is >20%. *Species present could be true shrubs, young trees and shrubs, or trees that are small or stunted due to environmental conditions.*
15	**Palustrine Emergent Wetland (Persistent)** – includes tidal and nontidal wetlands dominated by persistent emergent vascular plants, emergent mosses or lichens, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is <0.5%. Total vegetation cover is >80%. *Plants generally remain standing until the next growing season.*
Estuarine Wetlands	
16	**Estuarine Forested Wetland** – includes tidal wetlands dominated by woody vegetation ≥5 m in height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is ≥0.5%. Total vegetation coverage is >20%.
17	**Estuarine Scrub / Shrub Wetland** – includes tidal wetlands dominated by woody vegetation <5 m in height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is ≥0.5%. Total vegetation coverage is >20%.
18	**Estuarine Emergent Wetland** – Includes all tidal wetlands dominated by erect, rooted, herbaceous hydrophytes (excluding mosses and lichens). Wetlands that occur in tidal areas in which salinity due to ocean-derived salts is ≥0.5% and that are present for most of the growing season in most years. Total vegetation cover is >80%. *Perennial plants usually dominate these wetlands.*
Barren Land	
19	**Unconsolidated Shore** – includes material such as silt, sand, or gravel that is subject to inundation and redistribution due to the action of water. Substrates lack vegetation except for pioneering plants that become established during brief periods when growing conditions are favorable.
Water and Submerged Lands	
21	**Open Water** – include areas of open water, generally with <25% cover of vegetation or soil.
22	**Palustrine Aquatic Bed** – includes tidal and nontidal wetlands and deepwater habitats in which salinity due to ocean-derived salts is <0.5% and which are dominated by plants that grow and form a continuous cover principally on or at the surface of the water. These include algal mats, detached floating mats, and rooted vascular plant assemblages. Total vegetation cover is >80%.
23	**Estuarine Aquatic Bed** – includes tidal wetlands and deepwater habitats in which salinity due to ocean-derived salts is ≥0.5% and which are dominated by plants that grow and form a continuous cover principally on or at the surface of the water. These include algal mats, kelp beds, and rooted vascular plant assemblages. Total vegetation cover is >80%.

It is important to note that the NLCD classification scheme provides some detailed level II rural land cover information (e.g., three forest classes, two wetland classes, two agricultural classes). However, the classification scheme does not differentiate between residential, commercial and services, industrial, or transportation land cover as in the Anderson classification system. Rather, the emphasis in urbanized areas is on the identification of barren land and low-, medium-, and high-density development. This makes it somewhat difficult to compare NLCD urban/suburban land-cover information with more traditional land cover studies that differentiate between residential, commercial, industrial, and transportation land cover.

NOAA Coastal Change Analysis Program (C-CAP) Classification Scheme

NOAA's Coastal Change Analysis Program (C-CAP) produces a nationally standardized database of land-cover and land-change information for the coastal regions of the U.S (Table 9-4). In fact, C-CAP provides the "coastal expression" information for the NLCD

(NOAA C-CAP, 2014). C-CAP products provide inventories of coastal intertidal areas, wetlands, and adjacent uplands with the goal of monitoring these habitats by updating the land-cover maps every five years. C-CAP products are developed using multiple dates of remotely sensed imagery and consist of raster-based land-cover maps for each date of analysis, as well as a file that highlights what changes have occurred between these dates and where the changes were located (NOAA C-CAP, 2014).

The NOAA C-CAP classification system includes many of the same classes as the NLCD but provides very detailed information about wetland-related phenomena, especially in the coastal zone. Some of the class definitions come from the Cowardin wetland classification system discussed below. NOAA provides an online *C-CAP Land Cover Atlas* where the results of C-CAP land-cover mapping and change detection can be viewed (http://www.csc.noaa.gov/ccapatlas/).

U.S. Department of the Interior Fish & Wildlife Service *Classification of Wetlands and Deepwater Habitats of the United States*

The conterminous U.S. continues to lose inland and coastal wetland to agricultural, residential, and commercial land-use development. Therefore, there is great interest in wetland mapping. Detenbeck (2002) and EPA (2008) review numerous methods used to classify wetlands and the characteristics of several wetland classification schemes.

In the U.S., the Department of the Interior Fish & Wildlife Service is responsible for mapping and inventorying wetland. Therefore, they developed a wetland classification system that incorporates information extracted from remote sensor data and *in situ* measurement (Cowardin et al., 1979). Commonly referred to as the *Cowardin System*, it describes ecological taxa, arranges them in a system useful to resource managers, and provides uniformity of concepts and terms. Wetlands are classified based on plant characteristics, soils, and frequency of flooding. Ecologically important areas of deep water, traditionally not considered wetlands, are included in the classification as deep-water habitats.

Five systems form the highest level of the classification hierarchy: marine, estuarine, riverine, lacustrine, and palustrine (Figure 9-5). Marine and estuarine systems each have two subsystems: subtidal and intertidal. The riverine system has four subsystems: tidal, lower perennial, upper perennial, and intermittent. The lacustrine has two, littoral and limnetic, and the palustrine has no subsystem. Within the subsystems, classes are based on substrate material and flooding regime or on vegetative life form. The same classes may appear under one or more of the systems or subsystems. The distinguishing features of the Riverine system are presented in Figure 9-6a. Use of the classification system to characterize an Estuarine salt marsh near Bluffton, SC, is shown in Figure 9-6b. It highlights the characteristics of Tall Creekside and Intermediate *Spartina alterniflora*. This salt-tolerant habitat dominates much of the estuarine habitat along the eastern seaboard.

The Cowardin system was the first nationally recognized wetland classification scheme. It was adopted as the National Vegetation Classification Standard for wetlands mapping and inventory by the Wetlands Subcommittee of the Federal Geographic Data Committee. The Cowardin wetland classification system is the most practical scheme to use if you are going to extract wetland information from remotely sensed data and share the information with others interested in wetland-related problems. However, it is important to note that there are several other wetland classification schemes available (Detenbeck, 2002; EPA, 2008; NOAA C-CAP, 2014).

U.S. *National Vegetation Classification Standard (NVCS)*

The Vegetation Subcommittee of the U.S. Federal Geographic Data Committee has endorsed the *National Vegetation Classification Standard* (NVCS), which produces uniform vegetation resource data at the national level (FGDC, 2014). Separate categories are provided for natural and cultural vegetation, consistent with many other vegetation and land-cover classification schemes:

- *Natural vegetation* is defined as vegetation where ecological processes primarily determine species and site characteristics; that is, vegetation comprised of a largely spontaneously growing set of plant species that are shaped by both site and biotic processes. Natural vegetation forms recognizable physiognomic and floristic groupings that can be related to ecological site features.
- *Cultural vegetation* is defined as vegetation with a distinctive structure, composition, and development determined or influenced by regular human activity. The distinctive physiognomy, floristics, and dependence on human activity for its persistence set cultural vegetation apart from natural and/or semi-natural vegetation.

The *National Vegetation Classification Standard* is a good place to begin if a scientist is interested in extracting detailed vegetation information from remote sensor data and placing it in an ecologically sound vegetation classification system.

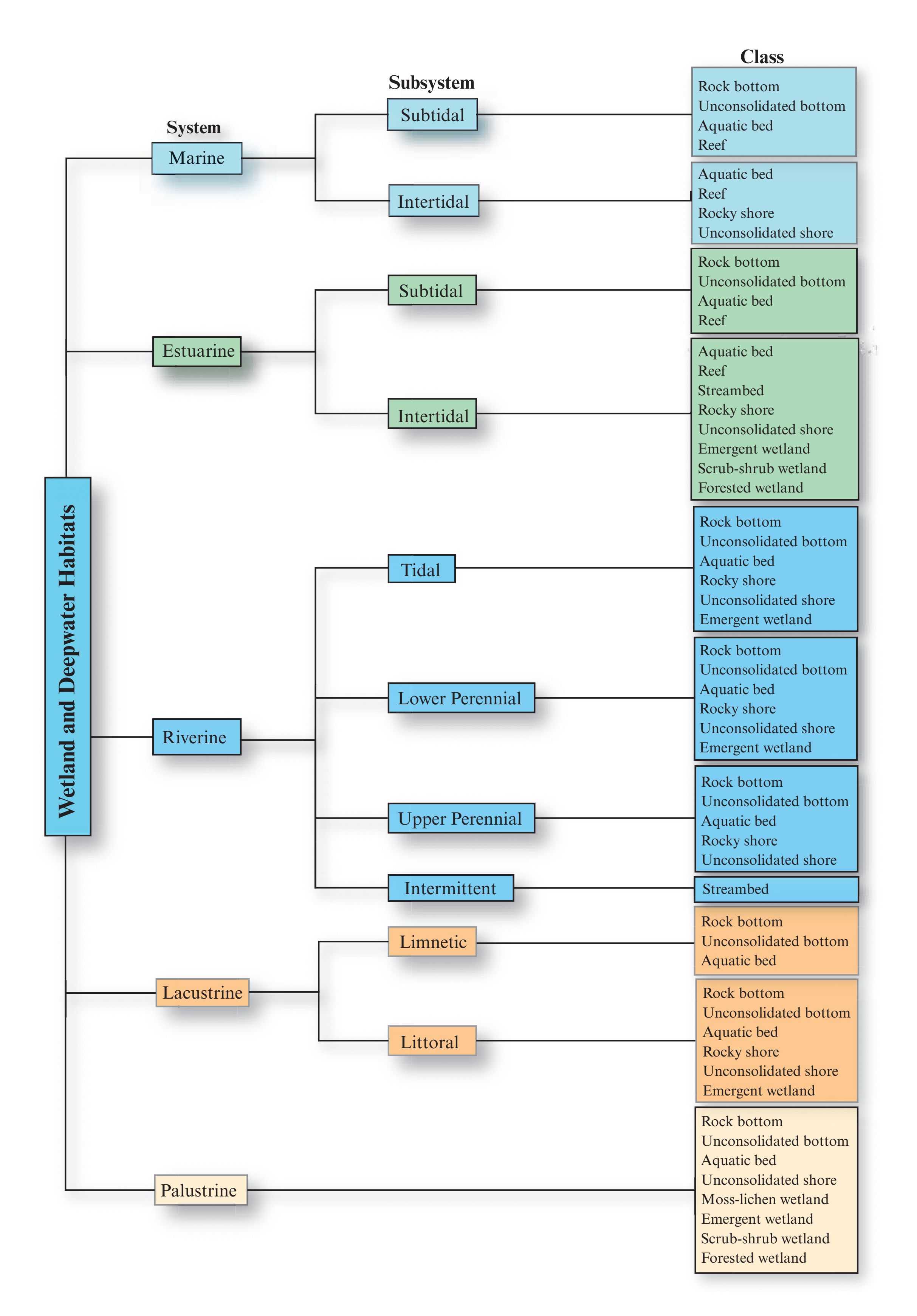
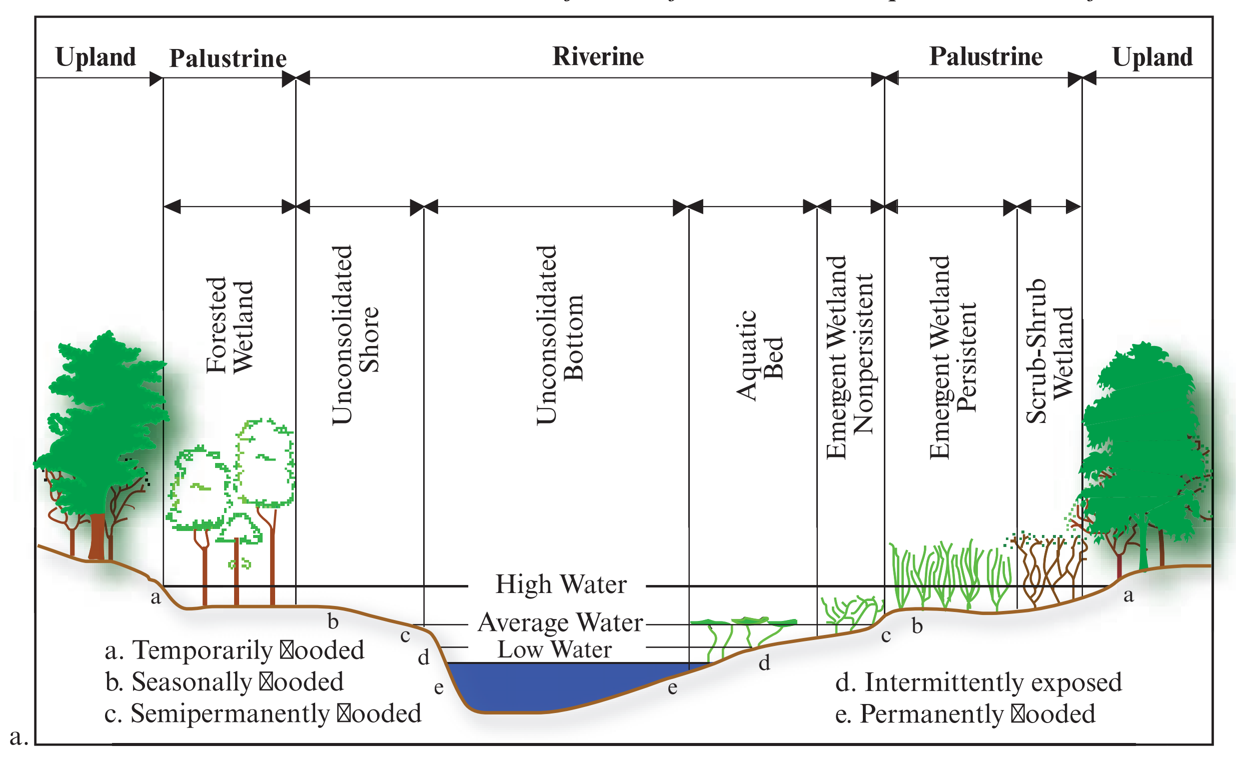
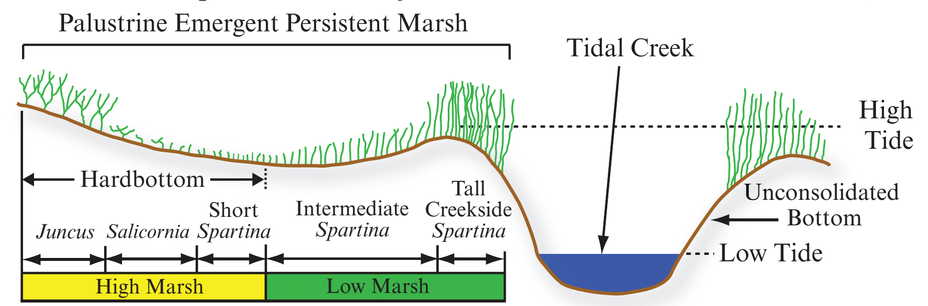


FIGURE 9–5 The *Classification of Wetlands and Deepwater Habitats of the United States* hierarchy of wetland habitat systems, subsystems, and classes (Cowardin et al., 1979). The palustrine system does not include deepwater habitats. The Cowardin system is the National Vegetation Classification Standard for wetlands mapping and inventory (FGDC Wetland Mapping Standard, 2014).

Cross-section of Riverine Habitats in the *Classification of Wetlands and Deepwater Habitats of the U.S.*



Cross-section of Estuarine Intertidal Habitats in the *Classification of Wetlands and Deepwater Habitats of the U.S.* for a Salt Marsh near Bluffton, SC



Tall Creekside *Spartina alterniflora* (Smooth Cordgrass) is more dense on the creekbanks where diurnal tidal flushing deposits nutrients and maintains reasonable levels of salinity. Over time, the creekside *Spartina* may trap suspended sediments, resulting in a gradual increase in the elevation of the tidal creek berm.



Intermediate *Spartina alterniflora* (Smooth Cordgrass) grows in a more stressed environment behind the creekbank, where nutrients are not as plentiful and there may be increased salinity levels in the substrate because of reduced tidal flushing.

FIGURE 9–6 a) Distinguishing features of habitats in the Riverine system of the *Classification of Wetlands and Deepwater Habitats of the United States* (Cowardin et al., 1979). b) Use of the classification scheme to characterize an Estuarine salt marsh near Bluffton, SC, along the May River (Jensen and Hodgson, 2011). A classification of this type of wetland habitat using object-based image analysis (OBIA) is presented in Figure 9-38c.

TABLE 9–5 *IGBP Land-Cover Classification System* modified for the creation of MODIS Land-Cover Type products (based on NASA Earth Observatory, 2002; Friedl et al., 2002, 2010; Giri et al., 2005).

Natural Vegetation	Description
Forest	
Evergreen Needleleaf	Dominated by woody vegetation with a cover >60% and height >2 m. Almost all trees remain green all year. Canopy is never without green foliage.
Evergreen Broadleaf	Dominated by woody vegetation with a cover >60% and height >2 m. Most trees and shrubs remain green year round. Canopy is never without green foliage.
Deciduous Needleleaf	Dominated by woody vegetation with a cover >60% and height >2 m. Seasonal needleleaf communities with an annual cycle of leaf-on and leaf-off periods.
Deciduous Broadleaf	Dominated by woody vegetation with a cover >60% and height >2 m. Broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.
Mixed Forest	Dominated by trees with a cover >60% and height >2 m. Communities with interspersed mixtures or mosaics of the other four forest types. None of the forest types exceeds 60% of landscape.
Shrubland, Grasslands, and Wetland	
Closed Shrublands	Woody vegetation <2 m tall and with shrub canopy cover >60%. The shrub foliage can be either evergreen or deciduous.
Open Shrublands	Woody vegetation <2 m tall and with shrub canopy cover between 10% and 60%. The shrub foliage can be either evergreen or deciduous.
Woody Savannas	Herbaceous and other understory systems, and with forest canopy cover between 30% and 60%. The forest cover height >2 m.
Savannas	Herbaceous and other understory systems, and with forest canopy cover between 10% and 30%. The forest cover height >2 m.
Grasslands	Lands with herbaceous types of cover. Tree and shrub cover is <10%.
Permanent Wetlands	Lands with a permanent mixture of water and herbaceous or woody vegetation. Vegetation can be present in salt, brackish, or fresh water.
Developed and Mosaic Lands	
Agriculture	
Croplands	Lands covered with temporary crops followed by harvest and a bare-soil period (e.g., single and multiple cropping systems). Perennial woody crops are classified as the appropriate forest or shrub land-cover type.
Cropland/Natural Vegetation Mosaic	
Cropland/Natural Vegetation Mosaic	Mosaic of croplands, forests, shrubland, and grasslands in which no one component comprises >60% of the landscape.
Urban	
Built-up	Land covered by buildings and other human-made structures.
Nonvegetated Lands	
Barren	
Barren or Sparsely Vegetated	Exposed soil, sand, rocks, or snow and ≤10% vegetated cover during the year.
Snow and Ice	Lands under snow/ice cover throughout the year.
Water	Oceans, seas, lakes, reservoirs, and rivers. Can be fresh or salt water.

International Geosphere-Biosphere Program *IGBP Land-Cover Classification System* Modified for the Creation of MODIS Land-Cover Type Products

If a scientist is interested in inventorying land cover at the regional, national, and global scale, then the modified International Geosphere-Biosphere Program *Land-Cover Classification System* may be appropriate. For example, Friedl et al. (2010) describe the MODIS Land Cover Type Products (MCD12Q1) that are produced each calendar year. These include:

- a 17-class IGBP land-cover classification;
- a 14-class University of Maryland classification;
- a 10-class LAI/FPAR (leaf-area-index/fraction of photosynthetically active radiation) classification;
- an 8-Biome classification; and a
- 12-class plant functional type classification.



FIGURE 9–7 a) Land-cover map of North America produced from *Terra* MODIS 1 × 1 km data collected between November 2000 and October 2001. b) Classification legend (images courtesy of MODIS Land Cover and Land Cover Dynamics group at the Boston University Center for Remote Sensing, NASA Goddard Space Flight Center (2004) and NASA Earth Observatory (2002)).

The characteristics of the IGBP classification scheme are found in Table 9-5. Collection 4 MODIS Global Land Cover Type (MLCT) products were produced at 1-km resolution. Collection 5 MLCT products are generated at 500-m spatial resolution. An example of Collection 4 MODIS land cover generated from images collected between November 2000 and October 2001 is shown in Figure 9-7.

Observations about Classification Schemes

Geographical information (including thematic maps and remote sensor data) is often imprecise. For example, there is usually a gradual transition at the interface

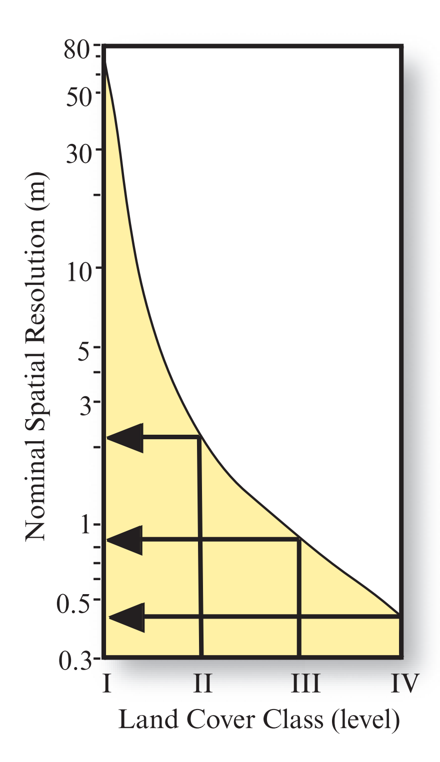


FIGURE 9–8 Nominal spatial resolution requirements as a function of the mapping requirements for Levels I to IV land-cover classes in the United States (based on Anderson et al., 1976). Note the dramatic increase in spatial resolution required to map Level II classes (based on Welch, 1982; Jensen and Cowen, 1999).

of forests and rangeland, yet many of the aforementioned classification schemes insist on a hard boundary between the classes at this transition zone. The schemes should contain fuzzy definitions because the thematic information they contain is fuzzy (Wang, 1990a). Fuzzy classification schemes are not currently standardized. They are typically developed by individual researchers for site-specific projects. The fuzzy classification systems developed may not be transferable to other environments. Therefore, we tend to see the use of existing hard classification schemes, which are rigid, based on *a priori* knowledge, and generally difficult to use. They continue to be widely employed because they are scientifically based and different individuals using the same classification system can compare results.

This brings us to another important consideration. If a reputable classification system already exists, it is foolish to develop an entirely new system that will probably be used only by ourselves. It is better to adopt or modify existing nationally or internationally recognized classification systems. This allows us to interpret the significance of our classification results in light of other studies and makes it easier to share data.

Finally, it should be reiterated that there is a general relationship between the level of detail in a classification scheme and the spatial resolution of remote sensor systems used to provide information. Welch (1982) summarized this relationship for mapping urban and suburban land use and land cover in the United States (Figure 9-8). A similar relationship exists when mapping vegetation (Botkin et al., 1984). For example, the sensor systems and spatial resolutions useful for discriminating vegetation from a global to an *in situ* perspective are summarized in Figure 9-9. This suggests that the level of detail in the desired classification system dictates the spatial resolution of the remote sensor data that should be used. Of course, the spectral resolution of the remote sensing system is also an important consideration, especially when inventorying vegetation, water, ice, snow, soil, and rock.

Training Site Selection and Statistics Extraction

An image analyst may select **training sites** within the image that are representative of the land-cover or land-use classes of interest after the classification scheme is adopted (Friedl et al., 2010). The training data should be of value if the environment from which they were obtained is relatively homogeneous. For example, if all the soils in a grassland region are composed of well-drained sandy-loam soil, then it is likely that grassland training data collected throughout the region would be representative. However, if the soil conditions change across the study area (e.g., one-half of the region has a perched water table with moist near-surface soil), it is likely that grassland training data acquired in the dry-soil part of the study area will *not* be representative of the spectral conditions for grassland found in the moist-soil portion of the study area. Thus, we have a *geographic signature extension* problem, meaning that it may not be possible to extend our grassland remote sensing training data through x, y space.

The easiest way to remedy this situation is to apply *geographical stratification* during the preliminary stages of a project. At this time all significant environmental factors that contribute to geographic signature extension problems should be identified, such as differences in soil type, water turbidity, crop species (e.g., two strains of wheat), unusual soil moisture conditions possibly caused by a thunderstorm that did not uniformly deposit its precipitation, scattered patches of atmospheric haze, and so on. Such environmental conditions should be carefully annotated on the imagery and the selection of training sites made based on the geographic stratification of these data. In such cases, it may be necessary to train the classifier over relatively short geographic distances. Each individual stratum may have to be classified separately. The final classification map of the entire region will then be a composite of the individual stratum classifications.

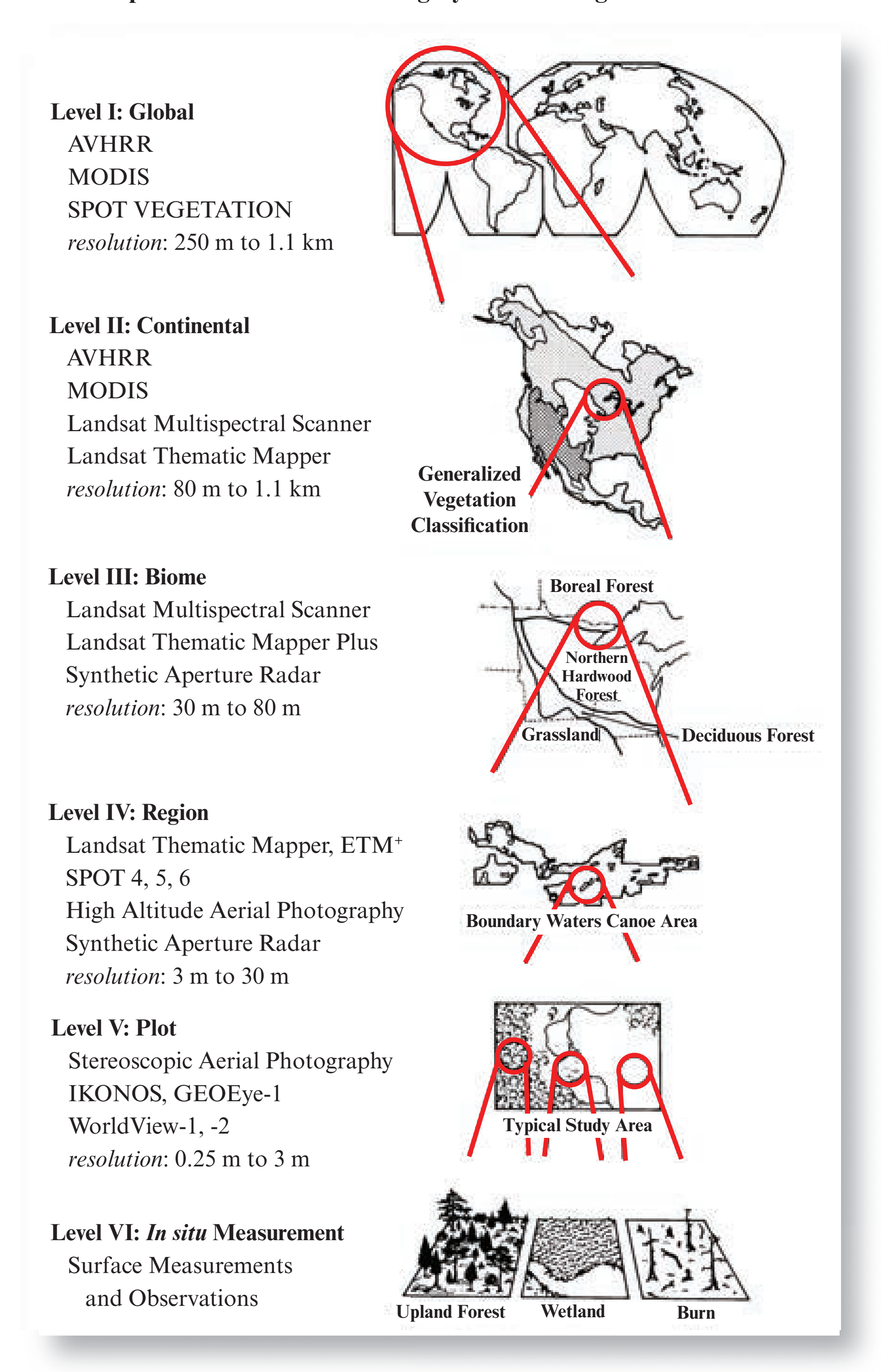


FIGURE 9–9 Relationship between the level of detail required and the spatial resolution of representative remote sensing systems for vegetation inventories.

However, if environmental conditions are homogeneous or can be held constant (e.g., through band ratioing or atmospheric correction), it may be possible to extend signatures vast distances in space, significantly reducing the training cost and effort. Additional research is required before the concept of geographic and temporal (through time) signature extension is fully understood (e.g., Jia et al., 2014).

Once spatial and temporal signature extension factors have been considered, the analyst selects representative training sites for each class and collects the spectral

statistics for each pixel found within each training site. Each site is usually composed of many pixels. The general rule is that if training data are being extracted from n bands, then a minimum of $>10n$ pixels of training data are collected for each class. This is sufficient to compute the variance–covariance matrices required by some classification algorithms.

There are a number of ways to collect the *training site* data, including:

- collection of *in situ* information such as tree type, height, percent canopy closure, and diameter-at-breast-height (dbh) measurements,
- on-screen selection of training data, and/or
- on-screen seeding of training data.

Ideally, each training site is visited in the field and its perimeter and/or centroid coordinates are measured directly using a global positioning system (GPS). Because the U.S. government has *selective availability* "off," the differentially corrected GPS *x, y* coordinates should be within – 1 m of their planimetric position. The GPS-derived *x, y* coordinates of a training site (e.g., a polygon enclosing a stand of oak) may then be input directly to the image processing system and used to extract training class (e.g., oak forest) statistics.

The analyst may also view the image on the computer screen and select polygonal areas of interest (AOI) (e.g., a stand of oak forest). Most image processing systems use a "rubber band" tool that allows the analyst to identify detailed AOIs. Conversely, the analyst may seed a specific location in the image using the cursor. The seed program begins at a single *x, y* location and evaluates neighboring pixel values in all bands of interest. Using criteria specified by the analyst, the seed algorithm expands outward like an amoeba as long as it finds pixels with spectral characteristics similar to the original seed pixel. This is a very effective way of collecting homogeneous training information.

Each pixel in each training site associated with a particular class (c) is represented by a *measurement vector,* X_c:

$$X_c = \begin{bmatrix} BV_{ij1} \\ BV_{ij2} \\ BV_{ij3} \\ \vdots \\ BV_{ijk} \end{bmatrix} \tag{9.1}$$

where $BV_{i,j,k}$ is the brightness value for the *i,j*th pixel in band k. The brightness values for each pixel in each band in each training class can then be analyzed statistically to yield a mean measurement vector, M_c, for each class:

$$M_c = \begin{bmatrix} \mu_{c1} \\ \mu_{c2} \\ \mu_{c3} \\ \vdots \\ \mu_{ck} \end{bmatrix} \tag{9.2}$$

where μ_{ck} represents the mean value of the data obtained for class c in band k. The raw measurement vector can also be analyzed to yield the covariance matrix for each class c:

$$V_c = V_{ckl} = \begin{bmatrix} Cov_{c11} & Cov_{c12} & \dots & Cov_{c1n} \\ Cov_{c21} & Cov_{c22} & \dots & Cov_{c2n} \\ \vdots & \vdots & \vdots & \vdots \\ Cov_{cn1} & Cov_{cn2} & \dots & Cov_{cnn} \end{bmatrix} \tag{9.3}$$

where Cov_{ckl} is the covariance of class c between bands k through l. For brevity, the notation for the covariance matrix for class c (i.e., V_{ckl}) will be shortened to just V_c. The same will be true for the covariance matrix of class d (i.e., $V_{dkl} = V_d$).

The mean, standard deviation, variance, minimum value, maximum value, variance–covariance matrix, and correlation matrix for the training statistics of five Charleston, SC, land-cover classes (residential, commercial, wetland, forest, and water) are listed in Table 9-6. These represent fundamental information on the spectral characteristics of the five classes.

Sometimes the manual selection of polygons results in the collection of training data with multiple modes in a training class histogram. This suggests that there are at least two types of land cover within the training area. This condition is not good when we are attempting to discriminate among individual classes. Therefore, it is a good practice to discard multimodal training data and retrain on specific parts of the polygon of interest until unimodal histograms are derived per class.

Positive spatial *autocorrelation* exists among pixels that are contiguous or close together. This means that adjacent pixels have a high probability of having similar brightness values. Training data collected from autocorrelated data tend to have reduced variance, which may be caused more by the way the sensor is collecting the data than by actual field conditions (e.g., most detectors dwell on an individual pixel for a very short time and may smear spectral information from one pixel to an adjacent pixel). The ideal situation is to collect training data within a region using every nth pixel or some other sampling criterion. The goal is to get

TABLE 9–6 Univariate and multivariate training statistics for the five land-cover classes using six bands of Landsat Thematic Mapper data of Charleston, SC.

a. Statistics for Residential

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Univariate Statistics						
Mean	70.6	28.8	29.8	36.7	55.7	28.2
Std. dev.	6.90	3.96	5.65	4.53	10.72	6.70
Variance	47.6	15.7	31.9	20.6	114.9	44.9
Minimum	59	22	19	26	32	16
Maximum	91	41	45	52	84	48
Variance–Covariance Matrix						
1	47.65					
2	24.76	15.70				
3	35.71	20.34	31.91			
4	12.45	8.27	12.01	20.56		
5	34.71	23.79	38.81	22.30	114.89	
7	30.46	18.70	30.86	12.99	60.63	44.92
Correlation Matrix						
1	1.00					
2	0.91	1.00				
3	0.92	0.91	1.00			
4	0.40	0.46	0.47	1.00		
5	0.47	0.56	0.64	0.46	1.00	
7	0.66	0.70	0.82	0.43	0.84	1.00

b. Statistics for Commercial

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Univariate Statistics						
Mean	112.4	53.3	63.5	54.8	77.4	45.6
Std. dev.	5.77	4.55	3.95	3.88	11.16	7.56
Variance	33.3	20.7	15.6	15.0	124.6	57.2
Minimum	103	43	56	47	57	32
Maximum	124	59	72	62	98	57
Variance–Covariance Matrix						
1	33.29					
2	11.76	20.71				
3	19.13	11.42	15.61			

b. Statistics for Commercial (Continued)

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
4	19.60	12.77	14.26	15.03		
5	–16.62	15.84	2.39	0.94	124.63	
7	–4.58	17.15	6.94	5.76	68.81	57.16
Correlation Matrix						
1	1.00					
2	0.45	1.00				
3	0.84	0.64	1.00			
4	0.88	0.72	0.93	1.00		
5	–0.26	0.31	0.05	0.02	1.00	
7	–0.10	0.50	0.23	0.20	0.82	1.00

c. Statistics for Wetland

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Univariate Statistics						
Mean	59.0	21.6	19.7	20.2	28.2	12.2
Std. dev.	1.61	0.71	0.80	1.88	4.31	1.60
Variance	2.6	0.5	0.6	3.5	18.6	2.6
Minimum	54	20	18	17	20	9
Maximum	63	25	21	25	35	16
Variance–Covariance Matrix						
1	2.59					
2	0.14	0.50				
3	0.22	0.15	0.63			
4	–0.64	0.17	0.60	3.54		
5	–1.20	0.28	0.93	5.93	18.61	
7	–0.32	0.17	0.40	1.72	4.53	2.55
Correlation Matrix						
1	1.00					
2	0.12	1.00				
3	0.17	0.26	1.00			
4	–0.21	0.12	0.40	1.00		
5	–0.17	0.09	0.27	0.73	1.00	
7	–0.13	0.15	0.32	0.57	0.66	1.00

d. Statistics for Forest

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Univariate Statistics						
Mean	57.5	21.7	19.0	39.1	35.5	12.5
Std. dev.	2.21	1.39	1.40	5.11	6.41	2.97
Variance	4.9	1.9	1.9	26.1	41.1	8.8
Minimum	53	20	17	25	22	8
Maximum	63	28	24	48	54	22
Variance–Covariance Matrix						
1	4.89					
2	1.91	1.93				
3	2.05	1.54	1.95			
4	5.29	3.95	4.06	26.08		
5	9.89	5.30	5.66	13.80	41.13	
7	4.63	2.34	2.22	3.22	16.59	8.84
Correlation Matrix						
1	1.00					
2	0.62	1.00				
3	0.66	0.80	1.00			
4	0.47	0.56	0.57	1.00		
5	0.70	0.59	0.63	0.42	1.00	
7	0.70	0.57	0.53	0.21	0.87	1.00

e. Statistics for Water

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Univariate Statistics						
Mean	61.5	23.2	18.3	9.3	5.2	2.7
Std. dev.	1.31	0.66	0.72	0.56	0.71	1.01
Variance	1.7	0.4	0.5	0.3	0.5	1.0
Minimum	58	22	17	8	4	0
Maximum	65	25	20	10	7	5
Variance–Covariance Matrix						
1	1.72					
2	0.06	0.43				
3	0.12	0.19	0.51			
4	0.09	0.05	0.05	0.32		

e. Statistics for Water (Continued)

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
5	–0.26	–0.05	–0.11	–0.07	0.51	
7	–0.21	–0.05	–0.03	–0.07	0.05	1.03
Correlation Matrix						
1	1.00					
2	0.07	1.00				
3	0.13	0.40	1.00			
4	0.12	0.14	0.11	1.00		
5	–0.28	–0.10	–0.21	–0.17	1.00	
7	–0.16	–0.08	–0.04	–0.11	0.07	1.00

non-autocorrelated training data. Unfortunately, most digital image processing systems do not provide this option in training data–collection modules.

Selecting the Optimum Bands for Image Classification: Feature Selection

Once the training statistics have been systematically collected from each band for each class of interest, a judgment must be made to determine the bands that are most effective in discriminating each class from all others. This process is commonly called **feature selection** (Dalponte et al., 2009; Fuchs et al., 2009). The goal is to delete from the analysis the bands that provide redundant spectral information. In this way the **dimensionality** (i.e., the number of bands to be processed) in the dataset may be reduced. This process minimizes the cost of the digital image classification process (but should not affect the accuracy). Feature selection may involve both statistical and graphical analysis to determine the degree of between-class separability in the remote sensor training data (Lillesand et al., 2008). Using statistical methods, combinations of bands are normally ranked according to their potential ability to discriminate each class from all others using n bands at a time (Beauchemin and Fung, 2001; Lu and Weng, 2007; Fuchs et al., 2009).

One might ask, "Why use graphical methods of feature selection if statistical techniques provide information sufficient to select the most appropriate bands for classification?" The reason is simple. An analyst may base a decision solely on the statistic, yet never obtain a basic understanding of the spectral nature of the data being analyzed. In effect, without ever visualizing where the spectral measurements cluster in n-dimensional feature space, each new supervised classification finds the analyst beginning anew, relying totally on the abstract statistical analysis. Many practitioners of remote sensing are by necessity very graphically literate; that is, they can readily interpret maps and graphs. Therefore, a graphic display of the statistical data is useful and often necessary for a thorough analysis of multispectral training data and feature selection. Graphic feature selection methods have been developed for this purpose.

Graphic Methods of Feature Selection

Feature space plots in two dimensions depict the distribution of all the pixels in the scene using two bands at a time (Figure 9-10). Such plots are often used as a backdrop for the display of various graphic feature selection methods. A typical plot usually consists of a 256×256 matrix (0 to 255 in the x-axis and 0 to 255 in the y-axis), which is filled with values in the following manner. Let us suppose that the first pixel in the entire dataset has a brightness value of 50 in band 1 and a value of 30 in band 3. A value of 1 is placed at location 50, 30 in the feature space plot matrix. If the next pixel in the dataset also has brightness values of 50 and 30 in bands 1 and 3, the value of this cell in the feature space matrix is incremented by 1, becoming 2. This logic is applied to each pixel in the scene. The brighter the pixel is in the feature space plot display, the greater the number of pixels having the same values in the two bands of interest.

Feature space plots provide insight into the actual information content of the image and the degree of between-band correlation. For example, in Figure 9-10a it is obvious that Thematic Mapper bands 1 (blue) and 3 (red) are highly correlated and that atmospheric scattering in band 1 (blue) results in a significant shift of the brightness values down the x-axis. Conversely, plots of bands 2 (green) and 4 (near-infrared) and bands 3 (red) and 4 have a much greater distribution of pixels within the spectral space and some interesting

Two-dimensional Feature Space Plots

FIGURE 9–10 Two-dimensional feature space plots of four pairs of Landsat TM data of Charleston, SC. a) TM bands 1 and 3, b) TM bands 2 and 4, c) TM bands 3 and 4, and d) TM bands 4 and 5. The brighter a particular pixel is in the display, the more pixels within the scene having that unique combination of band values.

bright locations, which correspond with important land-cover types (Figures 9-10bc). Finally, the plot of bands 4 (near-infrared) and 5 (middle-infrared) shows exceptional dispersion throughout the spectral space and some very interesting bright locations (Figure 9-10d). For this reason, a spectral feature space plot of bands 4 and 5 will be used as a backdrop for the next graphical feature selection method.

Cospectral parallelepiped or *ellipse plots* in two-dimensional feature space provide useful visual between-class separability information. They are produced using the mean, μ_{ck}, and standard deviation, σ_{ck}, of training class statistics for each class c and band k. For example, the original training statistics for five Charleston, SC, land-cover classes are portrayed in this manner and draped over the feature space plot of TM bands 4 and 5 in Figure 9-11. The lower and upper limits of the two-dimensional parallelepipeds (rectangles) were obtained using the mean $\pm 1\sigma$ of each band for each class. If only band 4 data were used to classify the scene, there would be confusion between classes 1 and 4, and if only band 5 data were used, there would be confusion between classes 3 and 4. However, when band 4 and 5 data are used at the same time to classify the

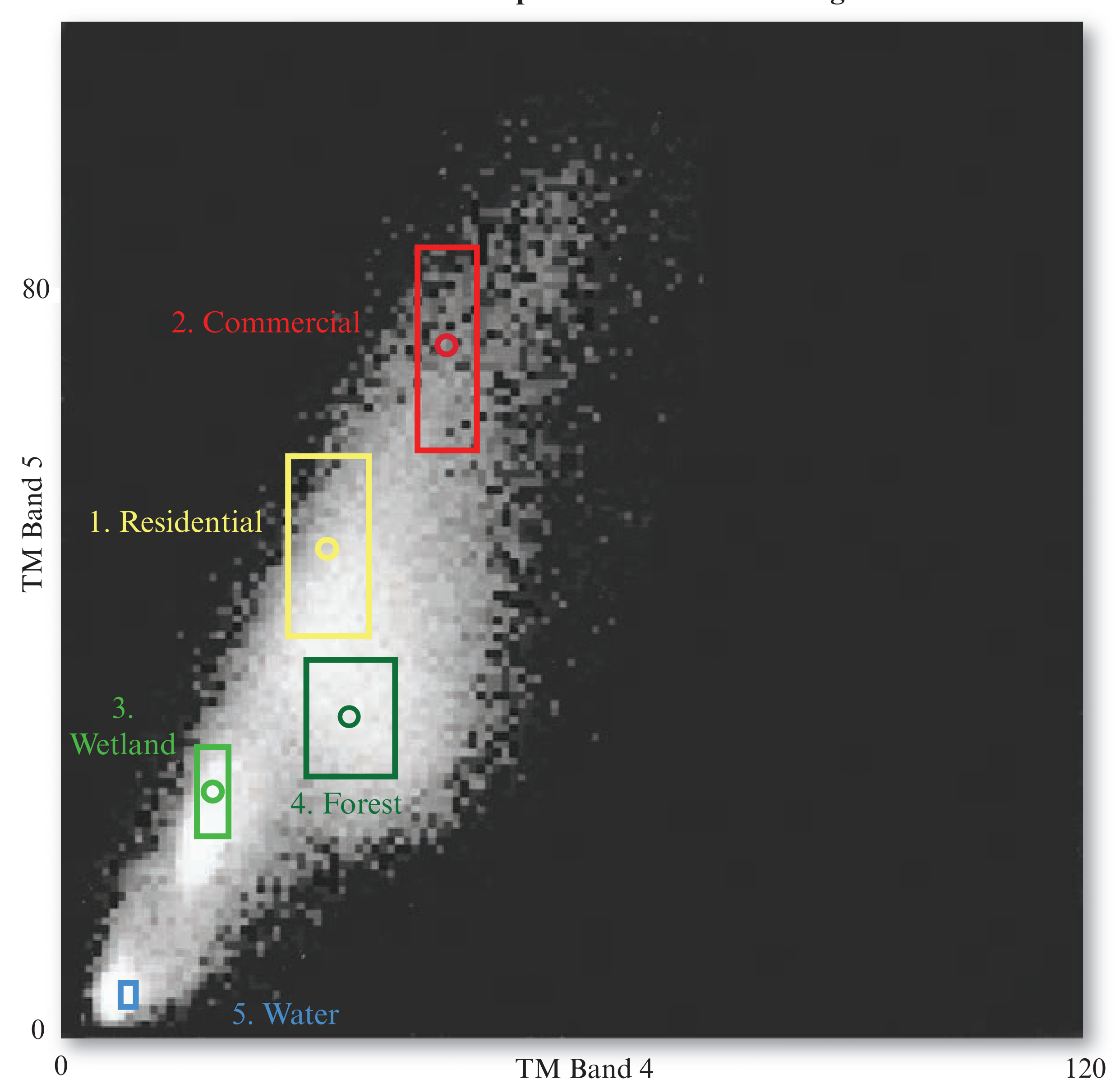


FIGURE 9–11 Plot of the Charleston, SC, Landsat TM training statistics for five classes measured in bands 4 and 5 displayed as cospectral parallelepipeds. The upper and lower limit of each parallelepiped is -1σ. The parallelepipeds are superimposed on a feature space plot of bands 4 and 5.

scene there appears to be good between-class separability among the five classes (at least at -1σ).

An evaluation of Figure 9-11 reveals that there are numerous water pixels in the scene found near the origin in bands 4 and 5. The water training class is located in this region. Similarly, the wetland training class is situated within the bright wetland region of band 4 and 5 spectral space. However, it appears that training data were not collected in the heart of the wetland region of spectral space. Such information is valuable because we may want to collect additional training data in the wetland region to see if we can capture more of the essence of the feature space. In fact, there may be two or more wetland classes residing in this portion of spectral space. Sophisticated digital image processing systems allow the analyst to select training data directly from this type of display, which contains 1) the training class parallelepipeds, and 2) the feature space plot. The analyst uses the cursor to interactively select training locations (they may be polygonal areas, not just parallelepipeds) within the feature space. If desired, these feature space partitions can be used as the actual decision logic during the classification phase of the project. This type of interactive feature space partitioning is very useful.

It is possible to display three bands of training data at once using *trispectral parallelepipeds* or *ellipses* in synthetic three-dimensional feature space (Figure 9-12). Jensen and Toll (1982) presented a method of displaying parallelepipeds in synthetic three-dimensional space and of interactively varying the viewpoint azimuth and elevation angles to enhance feature analysis and selection. Again, the mean, $⊠_{ck}$, and standard deviation, σ_{ck}, of training class statistics for each class c and band k were used to identify the lower and upper threshold values for each class and band. The analyst then selects a combination of three bands to portray because it is not possible to use all six bands at once in a three-dimensional display. Landsat TM bands 4, 5, and 7 are used in the following example; however, the method is applicable to any three-band subset. Each corner of a parallelepiped is identifiable by a unique set of x, y, z coordinates corresponding to either the lower or upper threshold value for the three bands under investigation (Figure 9-12).

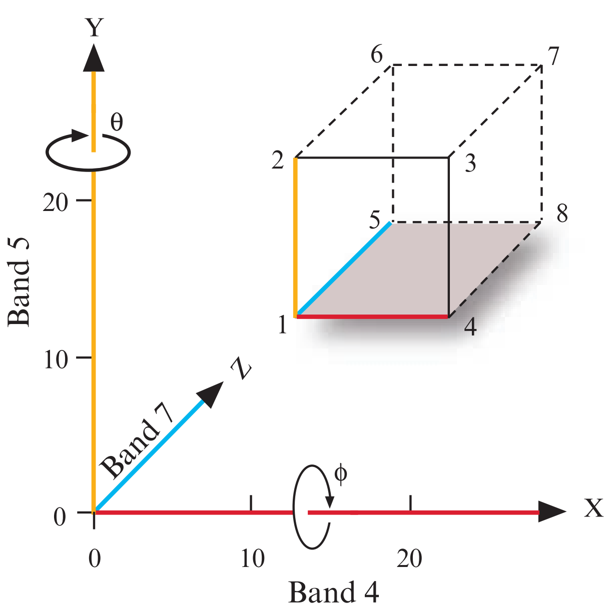


FIGURE 9–12 Simple parallelepiped displayed in pseudo three-dimensional space. Each of the eight corners represents a unique x, y, z coordinate corresponding to a lower or upper threshold value of the training data. For example, the original coordinates of point 4 are associated with 1) the upper threshold value of band 4, 2) the lower threshold value of band 5, and 3) the lower threshold value of band 7. The rotation matrix transformations cause the original coordinates to be rotated about the y-axis some θ radians, and the x-axis some ϕ radians.

The corners of the parallelepipeds may be viewed from a vantage point other than a simple frontal view of the x, y axes using three-dimensional coordinate transformation equations. The feature space may be rotated about any of the axes, although rotation around the x- and y-axes normally provides a sufficient number of viewpoints. Rotation about the x-axis ϕ radians and the y-axis θ radians is implemented using the following equations (Hodgson and Plews, 1989):

$$\underbrace{[X, Y, Z, 1]}_{P^{T'}} = \underbrace{[BVx, BVy, BVz, 1]}_{P^{T}} \bullet \tag{9.4}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi & 0 \\ 0 & \sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \bullet \begin{bmatrix} \cos\theta & 0 & \sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Negative signs of ϕ or θ are used for counterclockwise rotation and positive signs for clockwise rotation. This transformation causes the original brightness value coordinates, P^T, to be shifted about and contain specified depth information as vector $P^{T'}$. Display devices are two-dimensional (e.g., plotter surfaces or LCD screens); only the x and y elements of the transformed matrix $P^{T'}$ are used to draw the parallelepipeds.

Manipulation of the transformed coordinates associated with the Charleston, SC, training statistics is shown in Figure 9-13. All three of the bands (4, 5, and 7) are displayed in Figure 9-13a, except that the band 7 statistics are perpendicular (orthogonal) to the sheet of paper. By rotating the display 45⊠, the contribution of band 7 becomes apparent (Figure 9-13b). This represents a synthetic three-dimensional display of the parallelepipeds. As the display is rotated another 45⊠ to 90⊠, band 7 data collapse onto what was the band 4 axis (Figure 9-13c). The band 4 axis is now perpendicular to the page, just as band 7 was originally. The band 7, band 5 plot (Figure 9-13c) displays some overlap between wetland (3) and forest (4). By systematically specifying various azimuth and elevation angles, it is possible to display the parallelepipeds for optimal visual examination. This allows the analyst to obtain insight as to the consistent location of the training data in three-dimensional feature space.

In this example it is evident that just two bands, 4 and 5, provide as good if not better separation than all three bands used together. However, this may not be the very best set of two bands to use. It might be useful to evaluate other two- or three-band combinations. In fact, a certain combination of perhaps four or five bands used all at one time might be superior. The only way to determine if this is true is to perform statistical feature selection.

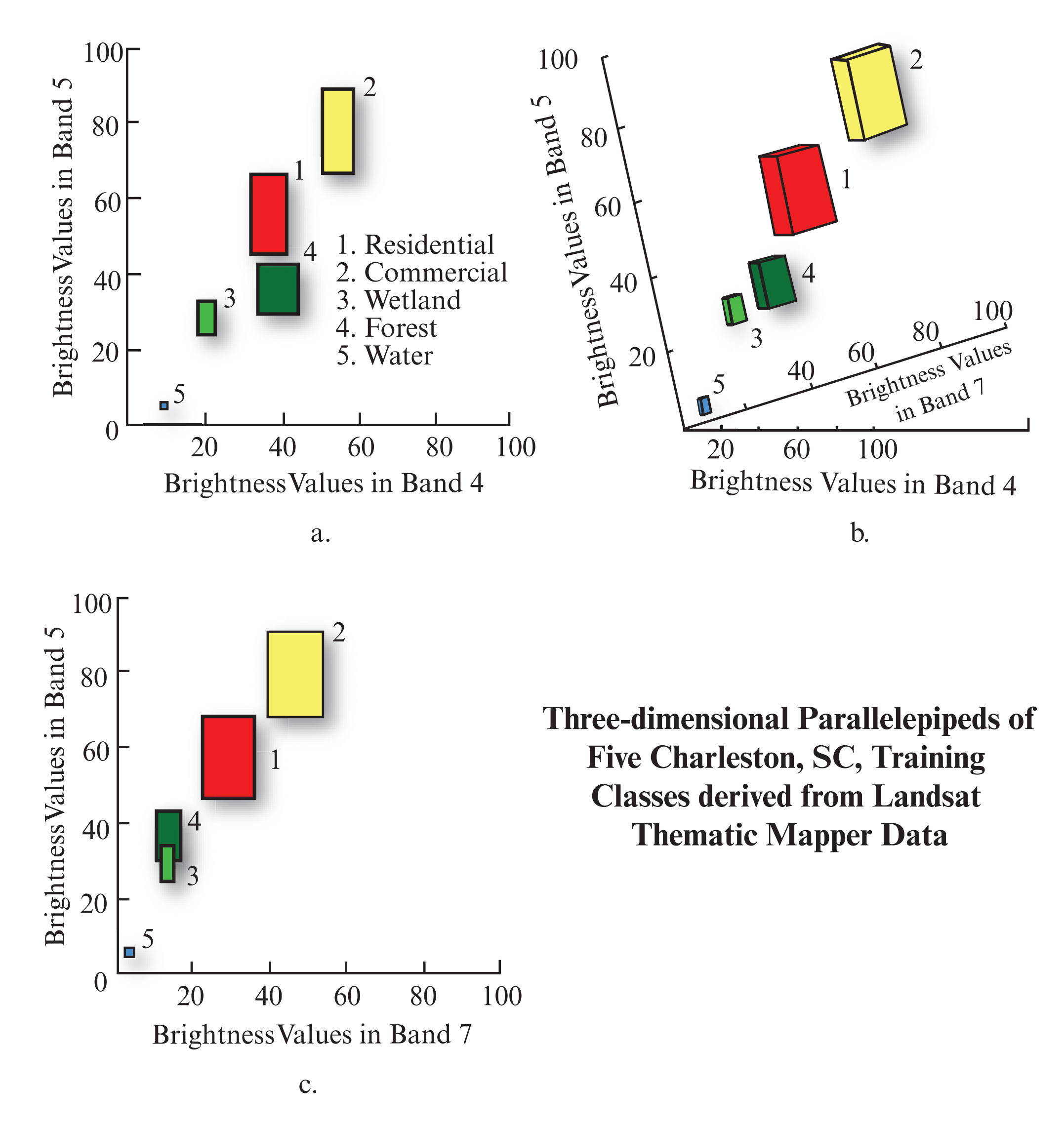


FIGURE 9–13 Development of the three-dimensional parallelepipeds of the five Charleston, SC, training classes derived from Landsat Thematic Mapper data. Only bands 4, 5, and 7 are used in this investigation. The data are rotated about the *y*-axis, 0°, 45°, 90°. At 0° and 90° (parts [a] and [c], respectively) we are actually looking at only two bands, analogous to the two-dimensional parallelepiped boxes shown in Figure 9-11. The third band lies perpendicular to the page we are viewing. Between such extremes, however, it is possible to obtain optimum viewing angles for visual analysis of training class statistics using three bands at once. Part (b) displays the five classes at a rotation of 45°, demonstrating that the classes are entirely separable using this three-band combination. However, it probably is not necessary to use all three bands since bands 4 and 5 alone will discriminate satisfactorily between the five classes, as shown in part (a). There would be a substantial amount of overlap between classes 3 and 4 if only bands 5 and 7 were used.

Statistical Methods of Feature Selection

Statistical methods of **feature selection** are used to quantitatively select which subset of bands (also referred to as features) provides the greatest degree of statistical separability between any two classes *c* and *d*. The basic problem of spectral pattern recognition is that given a spectral distribution of data in *n* bands of remotely sensed data, we must find a discrimination technique that will allow separation of the major land-cover categories with a minimum of error and a minimum number of bands. This problem is demonstrated diagrammatically using just one band and two classes in Figure 9-14. Generally, the more bands we analyze in a classification, the greater the cost and perhaps the greater the amount of redundant spectral information being used. When there is overlap, any decision rule that one could use to separate or distinguish between two classes must be concerned with two types of error (Figure 9-14):

1. A pixel may be assigned to a class to which it does not belong (an error of commission).

2. A pixel is not assigned to its appropriate class (an error of omission).

The goal is to select the optimum subset of bands and apply appropriate classification techniques to minimize both types of error in the classification process. If the training data for each class from each band are normally distributed, as suggested in Figure 9-14, it is possible to use either a transformed divergence or Jeffreys–Matusita distance equation to identify the optimum subset of bands to use in the classification procedure.

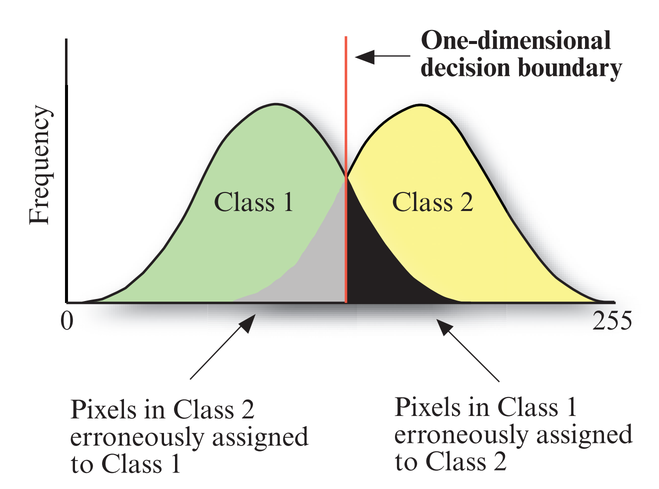


FIGURE 9–14 The basic problem in remote sensing pattern recognition classification is, given a spectral distribution of data in *n* bands (here just 1 band), to find an *n*-dimensional decision boundary that will allow the separation of the major classes (just 2 in this example) with a minimum of error and a minimum number of bands being evaluated. The dark areas of both distributions identify potential classification error.

Divergence *Divergence* was one of the first measures of statistical separability used in the machine processing of remote sensor data, and it is still widely used as a method of feature selection (e.g., Lu and Weng, 2007; Dalponte et al., 2009; Exelis ENVI, 2013). It addresses the basic problem of deciding what is the best *q*-band subset of *n* bands for use in the supervised classification process. The number of combinations *C* of *n* bands taken *q* at a time is:

$$C\binom{n}{q} = \frac{n!}{q!(n-q)!}. \qquad (9.5)$$

Thus, if there are six TM bands and we are interested in the three best bands to use in the classification of the Charleston scene, this results in 20 combinations that must be evaluated:

$$C\binom{6}{3} = \frac{6!}{3!(6-3)!} = \frac{720}{6(6)}$$

$$= 20 \text{ combinations.}$$

If the best two-band combination was desired, it would be necessary to evaluate 15 possible combinations.

Divergence is computed using the mean and covariance matrices of the class statistics collected in the training phase of the supervised classification. We will initiate the discussion by concerning ourselves with the statistical separability between just two classes, *c* and *d*. The degree of divergence or separability between *c* and *d*, Diver_{cd}, is computed according to the formula (Mausel et al., 1990):

$$\text{Diver}_{cd} = \frac{1}{2}tr[(V_c - V_d)(V_d^{-1} - V_c^{-1})] \qquad (9.6)$$
$$+ \frac{1}{2}tr[(V_c^{-1} + V_d^{-1})(M_c - M_d)(M_c - M_d)^T]$$

where *tr* [ ] is the trace of a matrix (i.e., the sum of the diagonal elements), V_c and V_d are the covariance matrices for the two classes under investigation, *c* and *d*, and M_c and M_d are the mean vectors for classes *c* and *d* (Konecny, 2014). It should be remembered that the sizes of the covariance matrices V_c and V_d are a function of the number of bands used in the training process (i.e., if six bands were trained upon, then both V_c and V_d would be matrices 6 × 6 in dimension). Divergence in this case would be used to identify the statistical separability of the two training classes using six bands of training data. However, this is not the usual goal of applying divergence. What we actually want to know is the optimum subset of *q* bands. For example, if *q* = 3, what subset of three bands provides the best separation between these two classes? Therefore, in our example, we would proceed to systematically apply the algorithm to the 20 three-band combinations, computing the divergence for our two classes of interest and eventually identifying the subset of bands, perhaps bands 2, 3, and 6, that results in the largest divergence value.

But what about the case where there are more than two classes? In this instance, the most common solution is to compute the *average divergence*, Diver_{avg}. This involves computing the average over all possible pairs of classes *c* and *d*, while holding the subset of bands *q* constant. Then, another subset of bands *q* is selected for the *m* classes and analyzed. The subset of features (bands) having the maximum average divergence may be the superior set of bands to use in the classification algorithm. This can be expressed:

$$\text{Diver}_{avg} = \frac{\sum_{c=1}^{m-1} \sum_{d=c+1}^{m} \text{Diver}_{cd}}{C}. \tag{9.7}$$

Using this, the band subset q with the highest average divergence would be selected as the most appropriate set of bands for classifying the m classes.

Unfortunately, outlying easily separable classes will weight average divergence upward in a misleading fashion to the extent that suboptimal feature subsets might be indicated as best (Ozkan and Erbek, 2003; Richards, 2013). Therefore, it is necessary to compute *transformed divergence*, TDiver_{cd}, expressed as:

$$\text{TDiver}_{cd} = 2000\left[1 - \exp\left(\frac{-\text{Diver}_{cd}}{8}\right)\right]. \tag{9.8}$$

This statistic gives an exponentially decreasing weight to increasing distances between the classes. It also scales the divergence values to lie between 0 and 2,000. For example, Table 9-7 demonstrates which bands are most useful when taken 1, 2, 3, 4, or 5 at a time. There is no need to compute the divergence using all six bands since this represents the totality of the dataset. It is useful, however, to calculate divergence with individual channels ($q = 1$), since a single channel might adequately discriminate among all classes of interest.

A transformed divergence value of 2,000 suggests excellent between-class separation. Above 1,900 provides good separation, while below 1,700 is poor. It can be seen that for the Charleston study, using any single band (Table 9-7a) would not produce results as acceptable as using bands 3 (red) and 4 (near-infrared) together (Table 9-7b). Several three-band combinations should yield good between-class separation for all classes. Most of them understandably include TM bands 3 and 4. But why should we use three, four, five, or six bands in the classification when divergence statistics suggest that very good between-class separation is possible using just two bands? We probably should not if the dimensionality of the dataset can be reduced by a factor of 3 (from 6 to 2) and classification results appear promising using just the two bands.

Bhattacharyya Distance Other methods of feature selection are also based on determining the separability between two classes at a time. For example, the *Bhattacharyya distance* assumes that the two classes c and d are Gaussian and that the means M_c and M_d and covariance matrices V_c and V_d are available. The algorithm is (Ifarraguerri and Prairie, 2004; Lu and Weng, 2007; Dalponte et al., 2009):

$$\text{Bhat}_{cd} = \frac{1}{8}(M_c - M_d)^T\left(\frac{V_c + V_d}{2}\right)^{-1}(M_c - M_d) \tag{9.9}$$

$$+ \frac{1}{2}\log_e\left[\frac{\left|\frac{V_c + V_d}{2}\right|}{\sqrt{(|V_c| \cdot |V_d|)}}\right].$$

To select the best q features (i.e., combination of bands) from the original n bands in an m-class problem, the Bhattacharyya distance is calculated between each $m(m-1)/2$ pair of classes for each possible way of choosing q features from n dimensions. The best q features are those dimensions whose sum of the Bhattacharyya distance between the $m(m-1)/2$ classes is highest (Konecny, 2014).

Jeffreys–Matusita Distance A saturating transformation applied to the Bhattacharyya distance measure yields the *Jeffreys–Matusita distance* (often referred to as the JM distance) computed using (Ferro and Warner, 2002; Ifarraguerri and Prairie, 2004; Dalponte et al., 2009):

$$\text{JM}_{cd} = \sqrt{2(1 - e^{-\text{Bhat}_{cd}})}. \tag{9.10}$$

The JM distance has a saturating behavior with increasing class separation like transformed divergence. However, it is not as computationally efficient as transformed divergence.

Ifarraguerri and Prairie (2004) developed a feature selection method that built on the Jeffreys–Matusita distance measure. They counted the number of times that each band was selected as the best when used in combination with other bands (i.e., 2 bands at a time, 3 bands at a time, etc.), during the band selection process. They plotted the band counts as two-dimensional histograms. The relative heights of the histogram bins are the relative frequencies with which each band is selected. The graphic provides a good visual indication of the desirability of each band in maximizing the distance metric.

Correlation Matrix Feature Selection The *Correlation Matrix feature selection* methodology is useful when trying to determine the optimum bands when creating vegetation and other types of indices. For example, Im et al. (2012) used Correlation Matrix feature selection to identify the optimum hyperspectral bands in the region from 400 to 2,500 nm to predict leaf-area-index (LAI) on hazardous waste sites in Utah and Arizona. The goal was to identify the optimum bands to use in the generic vegetation index:

TABLE 9–7 Divergence statistics for the five Charleston, SC, land-cover classes evaluated using 1, 2, 3, 4, and 5 Thematic Mapper band combinations at one time.

Band Combinations	Average Divergence	Divergence (upper number) and Transformed Divergence (lower number)									
		Class Combinations (1 = residential; 2 = commercial; 3 = wetland; 4 = forest; 5 = water)									
		1	**1**	**1**	**1**	**2**	**2**	**2**	**3**	**3**	**4**
		2	**3**	**4**	**5**	**3**	**4**	**5**	**4**	**5**	**5**
a. One Band at a Time											
1	1583	45 1993	36 1977	23 1889	38 1982	600 2000	356 2000	803 2000	1 198	3 651	7 1145
2	1588	34 1970	67 2000	15 1786	54 1998	1036 2000	286 2000	1090 2000	1 246	5 988	5 890
3	1525	54 1998	107 2000	39 1985	160 2000	1591 2000	576 2000	2071 2000	1 286	3 642	1 339
4	1748	19 1809	47 1994	0 70	1238 2000	209 2000	13 1603	3357 2000	60 1999	210 2000	1466 2000
5	1636	4 779	26 1920	7 1194	2645 2000	77 2000	29 1947	5300 2000	2 523	556 2000	961 2000
7	1707	6 1061	61 1999	18 1795	345 2000	238 2000	74 2000	940 2000	1 213	63 1999	56 1998
b. Two Bands at a Time											
1 2	1709	51 1997	92 2000	26 1919	85 2000	1460 2000	410 2000	1752 2000	2 463	8 1256	10 1457
1 3	1709	56 1998	125 2000	40 1987	182 2000	1888 2000	589 2000	2564 2000	2 418	7 1196	11 1490
1 4	1996	55 1998	100 2000	32 1962	1251 2000	941 2000	446 2000	3799 2000	66 1999	219 2000	1525 2000
1 5	1896	54 1998	71 2000	28 1939	3072 2000	778 2000	497 2000	7838 2000	6 1029	585 2000	1038 2000
1 7	1852	52 1997	107 2000	28 1939	426 2000	944 2000	421 2000	2065 2000	3 586	63 1999	76 2000
2 3	1749	57 1998	140 2000	42 1990	170 2000	2099 2000	593 2000	2345 2000	2 524	13 1599	9 1382
2 4	1992	35 1976	103 2000	28 1941	1256 2000	1136 2000	356 2000	3985 2000	65 1999	228 2000	1529 2000
2 5	1856	35 1976	86 2000	20 1826	2795 2000	1068 2000	328 2000	6932 2000	4 760	560 2000	979 2000
2 7	1829	37 1980	111 2000	24 1902	423 2000	1148 2000	292 2000	2192 2000	2 405	69 2000	66 1999
3 4	2000	101 2000	124 2000	61 1999	1321 2000	1606 2000	905 2000	4837 2000	80 2000	210 2000	1487 2000
3 5	1895	59 1999	114 2000	45 1992	3206 2000	1609 2000	740 2000	9142 2000	5 964	597 2000	1024 2000
3 7	1845	63 1999	131 2000	41 1989	525 2000	1610 2000	606 2000	3122 2000	2 469	65 1999	59 1999

TABLE 9–7 Divergence statistics for the five Charleston, SC, land-cover classes evaluated using 1, 2, 3, 4, and 5 Thematic Mapper band combinations at one time. (Continued)

Band Combinations	Average Divergence	Divergence (upper number) and Transformed Divergence (lower number)									
		Class Combinations (1 = residential; 2 = commercial; 3 = wetland; 4 = forest; 5 = water)									
		1	1	1	1	2	2	2	3	3	4
		2	3	4	5	3	4	5	4	5	5
4 5	1930	21 1851	52 1997	11 1468	4616 2000	231 2000	37 1981	10376 2000	98 2000	889 2000	2902 2000
4 7	1970	20 1844	76 2000	21 1857	1742 2000	309 2000	79 2000	4740 2000	86 2000	285 2000	1599 2000
5 7	1795	6 1074	62 1999	24 1900	2870 2000	246 2000	97 2000	5956 2000	5 978	598 2000	989 2000
c. Three Bands at a Time											
1 2 3	1815	59 1999	154 2000	44 1992	191 2000	2340 2000	613 2000	2821 2000	3 643	16 1745	17 1774
1 2 4	1999	95 2000	142 2000	40 1986	1266 2000	1662 2000	675 2000	4381 2000	68 2000	236 2000	1573 2000
1 2 5	1909	58 1999	118 2000	32 1964	3201 2000	1564 2000	604 2000	9281 2000	7 1129	589 2000	1045 2000
1 2 7	1868	57 1998	146 2000	30 1953	493 2000	1653 2000	494 2000	3176 2000	4 732	69 2000	80 2000
1 3 4	2000	117 2000	150 2000	64 1999	1329 2000	1905 2000	985 2000	5120 2000	86 2000	219 2000	1534 2000
1 3 5	1920	60 1999	137 2000	51 1997	3569 2000	1902 2000	863 2000	11221 2000	7 1202	622 2000	1088 2000
1 3 7	1872	63 1999	157 2000	45 1993	580 2000	1935 2000	669 2000	3879 2000	4 731	66 1999	79 2000
1 4 5	1998	82 2000	105 2000	36 1979	4923 2000	978 2000	635 2000	12361 2000	104 2000	906 2000	2955 2000
1 4 7	1998	82 2000	129 2000	37 1980	1777 2000	1055 2000	610 2000	5452 2000	93 2000	288 2000	1669 2000
1 5 7	1924	56 1998	109 2000	37 1982	3405 2000	956 2000	508 2000	8948 2000	8 1261	627 2000	1077 2000
2 3 4	2000	117 2000	156 2000	63 1999	1331 2000	2119 2000	956 2000	4971 2000	81 2000	229 2000	1530 2000
2 3 5	1908	62 1999	147 2000	47 1994	3221 2000	2120 2000	749 2000	9480 2000	6 1082	605 2000	1034 2000
2 3 7	1865	66 1999	160 2000	46 1994	541 2000	2113 2000	617 2000	3480 2000	3 661	74 2000	69 2000
2 4 5	1994	38 1984	108 2000	31 1956	4674 2000	1158 2000	385 2000	11402 2000	103 2000	896 2000	2946 2000
2 4 7	1996	40 1986	125 2000	34 1970	1771 2000	1191 2000	367 2000	5511 2000	90 2000	300 2000	1668 2000
2 5 7	1906	38 1982	113 2000	33 1968	3050 2000	1157 2000	365 2000	7757 2000	7 1113	594 2000	1006 2000

TABLE 9–7 Divergence statistics for the five Charleston, SC, land-cover classes evaluated using 1, 2, 3, 4, and 5 Thematic Mapper band combinations at one time. (Continued)

Band Combinations	Average Divergence	Divergence (upper number) and Transformed Divergence (lower number)									
		Class Combinations (1 = residential; 2 = commercial; 3 = wetland; 4 = forest; 5 = water)									
		1	1	1	1	2	2	2	3	3	4
		2	3	4	5	3	4	5	4	5	5
3 4 5	2000	106 2000	129 2000	65 1999	5031 2000	1622 2000	1037 2000	13505 2000	120 2000	914 2000	2935 2000
3 4 7	2000	111 2000	144 2000	63 1999	1841 2000	1644 2000	955 2000	6309 2000	102 2000	285 2000	1626 2000
3 5 7	1927	66 1999	134 2000	63 1999	3453 2000	1648 2000	823 2000	9900 2000	8 1268	631 2000	1054 2000
4 5 7	1979	22 1870	83 2000	26 1923	5003 2000	362 2000	114 2000	11477 2000	105 2000	944 2000	2994 2000
d. Four Bands at a Time											
1 2 3 4	2000	167 2000	177 2000	65 1999	1339 2000	2361 2000	1151 2000	5259 2000	87 2000	238 2000	1575 2000
1 2 3 5	1929	63 1999	165 2000	54 1998	3582 2000	2355 2000	876 2000	11525 2000	8 1294	630 2000	1095 2000
1 2 3 7	1888	67 2000	182 2000	49 1996	595 2000	2369 2000	683 2000	4222 2000	5 885	75 2000	87 2000
1 2 4 5	1999	115 2000	147 2000	46 1994	4971 2000	1696 2000	901 2000	13287 2000	108 2000	913 2000	2987 2000
1 2 4 7	1999	110 2000	165 2000	45 1993	1801 2000	1731 2000	868 2000	6161 2000	96 2000	303 2000	1725 2000
1 2 5 7	1932	61 1999	148 2000	41 1989	3564 2000	1665 2000	614 2000	10579 2000	9 1331	633 2000	1085 2000
1 3 4 5	2000	133 2000	156 2000	74 2000	5293 2000	1931 2000	1283 2000	15187 2000	127 2000	928 2000	2976 2000
1 3 4 7	2000	134 2000	172 2000	69 2000	1863 2000	1955 2000	1184 2000	6814 2000	110 2000	289 2000	1682 2000
1 3 5 7	1940	66 2000	159 2000	66 2000	3919 2000	1954 2000	901 2000	12411 2000	10 1397	665 2000	1129 2000
1 4 5 7	1999	88 2000	135 2000	42 1990	5422 2000	1105 2000	659 2000	13950 2000	112 2000	970 2000	3068 2000
2 3 4 5	2000	122 2000	161 2000	67 2000	5040 2000	2133 2000	1093 2000	13663 2000	121 2000	933 2000	2981 2000
2 3 4 7	2000	132 2000	173 2000	65 1999	1848 2000	2143 2000	1023 2000	6509 2000	103 2000	302 2000	1670 2000
2 3 5 7	1937	69 2000	163 2000	68 2000	3476 2000	2144 2000	837 2000	10308 2000	9 1370	639 2000	1062 2000
2 4 5 7	1997	41 1987	131 2000	38 1983	5079 2000	1229 2000	397 2000	12641 2000	110 2000	951 2000	3037 2000
3 4 5 7	2000	112 2000	148 2000	74 2000	5436 2000	1665 2000	1066 2000	14688 2000	125 2000	971 2000	3030 2000

TABLE 9–7 Divergence statistics for the five Charleston, SC, land-cover classes evaluated using 1, 2, 3, 4, and 5 Thematic Mapper band combinations at one time. (Continued)

Band Combinations	Average Divergence	Divergence (upper number) and Transformed Divergence (lower number)									
		Class Combinations (1 = residential; 2 = commercial; 3 = wetland; 4 = forest; 5 = water)									
		1	1	1	1	2	2	2	3	3	4
		2	3	4	5	3	4	5	4	5	5
e. Five Bands at a Time											
1 2 3 4 5	2000	176 2000	183 2000	75 2000	5302 2000	2384 2000	1422 2000	15334 2000	128 2000	947 2000	3019 2000
1 2 3 4 7	2000	176 2000	196 2000	71 2000	1871 2000	2393 2000	1316 2000	7015 2000	111 2000	305 2000	1726 2000
1 2 3 5 7	1948	70 2000	184 2000	72 2000	3940 2000	2386 2000	919 2000	12798 2000	11 1479	673 2000	1135 2000
1 2 4 5 7	2000	117 2000	171 2000	50 1996	5487 2000	1770 2000	920 2000	15021 2000	115 2000	977 2000	3101 2000
1 3 4 5 7	2000	138 2000	176 2000	80 2000	5803 2000	1979 2000	1294 2000	16829 2000	132 2000	994 2000	3089 2000
2 3 4 5 7	2000	134 2000	177 2000	77 2000	5443 2000	2161 2000	1130 2000	14893 2000	126 2000	987 2000	3072 2000

$$VI2 = \frac{B_2 - B_1}{B_2 + B_1}$$

where B_1 and B_2 are bands in the region from 400 to 2,500 nm and B_1 is always smaller in value than B_2. The coefficient of determination (R^2) was computed between the VI2 under investigation and *in situ* LAI measurements for each of the possible 15,750 possible VI2 band combinations. The darker the shade of blue in Figure 9-15, the higher the correlation between the VI2 and LAI. The darker the shade of red, the lower the correlation between the VI2 and LAI. Mahlein et al. (2013) used a related feature extraction algorithm called RELIEF-F to identify the optimum bands for creating Sugarbeet rust and powdery mildew detection indices from hyperspectral data.

Fuchs et al. (2009) used stepwise linear regression modeling to select which original QuickBird and remote sensing–derived variables to use in their prediction of above ground carbon. The variables tested included original bands, Kauth-Thomas Tasseled Cap, Principal Components Analysis (PCA), numerous vegetation indices and texture measures.

Principal Components Analysis (PCA) Feature Selection As discussed in Chapter 8, principal components analysis (PCA) can be applied to a multispectral dataset to identify the principal components and the amount of variance explained by each component. The newly-created PCA images that account for the

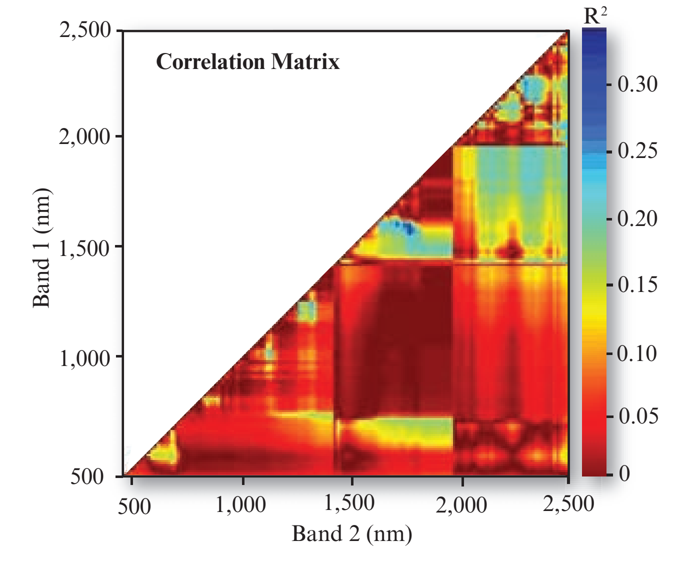


FIGURE 9–15 An example of the use of a Correlation Matrix feature (band) selection. The goal was to identify the optimum bands from 400 to 2,500 nm for use in the generic vegetation index VI2 = $(B_2–B_1) / (B_2+B_1)$ regressed against *in situ* LAI measurements. The higher the coefficient of determination (R^2) for each of the 15,750 possible VI2s regressed with *in situ* LAI measurements, the darker the shade of blue. The darker the shade of red, the lower the correlation between a VI2 and LAI (adapted from Im et al., 2012).

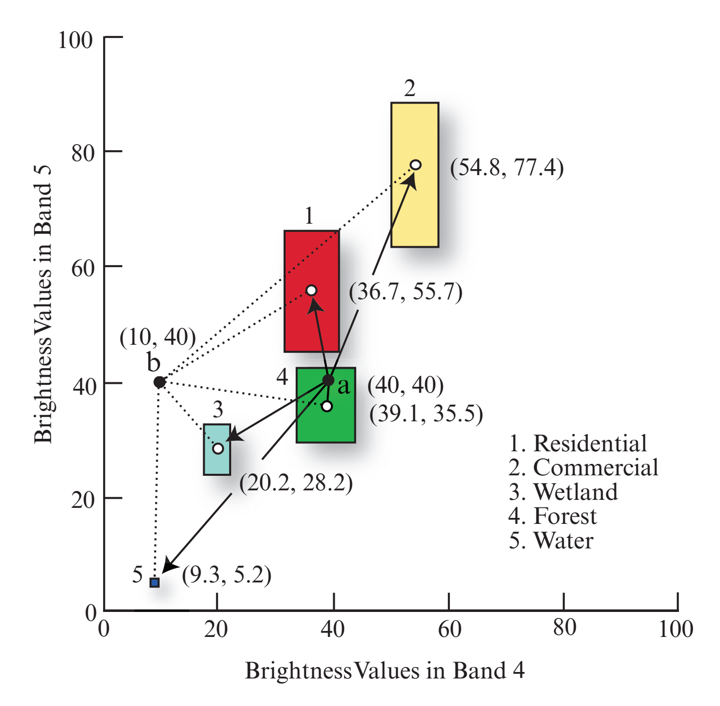


FIGURE 9–16 Points *a* and *b* are pixels in the image to be classified. Pixel *a* has a brightness value of 40 in band 4 and 40 in band 5. Pixel *b* has a brightness value of 10 in band 4 and 40 in band 5. The boxes represent the *parallelepiped* decision rule associated with a $\pm 1\sigma$ classification. The vectors (*arrows*) represent the distance from *a* and *b* to the mean of all classes in a *minimum distance to means* classification algorithm. Refer to Tables 9-8 and 9-9 for the results of classifying points *a* and *b* using both classification techniques.

most variance can then be used when collecting training data instead of the original remote sensor data. In this situation, PCA in effect performs feature selection and reduces the dimensionality of the dataset that is used for training and classification (Rocchini et al., 2007).

Select the Appropriate Classification Algorithm

Various supervised classification algorithms can be used to assign an unknown pixel to one of *m* possible classes (Jensen et al., 2009). The choice of a particular classifier or decision rule depends on the nature of the input data and the desired output and is very important. *Parametric* classification algorithms assume that the observed measurement vectors X_c obtained for each class in each spectral band during the training phase of the supervised classification are Gaussian; that is, they are normally distributed (Schowengerdt, 2007). Conversely, *nonparametric* classification algorithms make no such assumption (Lu and Weng, 2007).

Several widely adopted nonparametric classification algorithms include:

- one-dimensional *density slicing* (e.g., using a single band of imagery; please refer to Chapter 8),
- parallelepiped,
- minimum distance,
- nearest-neighbor, and
- neural network and expert system analysis (Chapter 10).

The most widely adopted parametric classification algorithms is:

- the maximum likelihood algorithm.

It is instructive to review the logic of several classification algorithms in this and subsequent chapters.

Parallelepiped Classification Algorithm

This digital image classification decision rule is based on simple Boolean "and/or" logic. Training data in *n* spectral bands are used to perform the classification. Along with other algorithms, it is used in sophisticated object-based image analysis (OBIA) programs such as *e*Cognition (Trimble, 2014). Brightness values from each pixel of the multispectral imagery are used to produce an *n*-dimensional mean vector, $M_c = (\mu_{ck}, \mu_{c2}, \mu_{c3}, \ldots, \mu_{cn})$ with μ_{ck} being the mean value of the training data obtained for class *c* in band *k* out of *m* possible classes, as previously defined. σ_{ck} is the standard deviation of the training data class *c* of band *k* out of *m* possible classes. In this discussion we will evaluate all five Charleston classes using just bands 4 and 5 of the training data.

TABLE 9–8 Example of parallelepiped classification logic for pixels *a* and *b* in Figure 9-16.

Class	Lower Threshold, L_{ck}	Upper Threshold, H_{ck}	Does pixel *a* (40, 40) satisfy criteria for this class in this band? $L_{ck} \leq a \leq H_{ck}$	Does pixel *b* (10, 40) satisfy criteria for this class in this band? $L_{ck} \leq b \leq H_{ck}$
1.Residential				
Band 4	36.7 – 4.53 = 31.27	36.7 + 4.53 = 41.23	Yes	No
Band 5	55.7 – 10.72 = 44.98	55.7 + 10.72 = 66.42	No	No
2.Commercial				
Band 4	54.8 – 3.88 = 50.92	54.8 + 3.88 = 58.68	No	No
Band 5	77.4 – 11.16 = 66.24	77.4 + 11.16 = 88.56	No	No
3.Wetland				
Band 4	20.2 – 1.88 = 18.32	20.2 + 1.88 = 22.08	No	No
Band 5	28.2 – 4.31 = 23.89	28.2 + 4.31 = 32.51	No	No
4.Forest				
Band 4	39.1 – 5.11 = 33.99	39.1 + 5.11 = 44.21	Yes	No
Band 5	35.5 – 6.41 = 29.09	35.5 + 6.41 = 41.91	Yes, assign pixel to class 4, forest. STOP.	No
5.Water				
Band 4	9.3 – 0.56 = 8.74	9.3 + 0.56 = 9.86	—	No
Band 5	5.2 – 0.71 = 4.49	5.2 + 0.71 = 5.91	—	No, assign pixel to unclassified category. STOP.

Using a one-standard-deviation threshold (as shown in Figure 9-16), a parallelepiped algorithm decides BV_{ijk} is in class c if, and only if:

$$\mu_{ck} - \sigma_{ck} \leq BV_{ijk} \leq \mu_{ck} + \sigma_{ck} \qquad (9.11)$$

where

$c = 1, 2, 3, \ldots, m,$ number of classes, and
$k = 1, 2, 3, \ldots, n,$ number of bands.

Therefore, if the low and high decision boundaries are defined as

$$L_{ck} = \mu_{ck} - \sigma_{ck} \qquad (9.12)$$

and

$$H_{ck} = \mu_{ck} + \sigma_{ck}, \qquad (9.13)$$

the parallelepiped algorithm becomes

$$L_{ck} \leq BV_{ijk} \leq H_{ck}. \qquad (9.14)$$

These decision boundaries form an n-dimensional parallelepiped in feature space. If the pixel value lies above the low threshold and below the high threshold for all n bands evaluated, it is assigned to that class (see point *a* in Figure 9-16). When an unknown pixel does not satisfy any of the Boolean logic criteria (point *b* in Figure 9-16), it is assigned to an "unclassified" category. Although it is only possible to analyze visually up to three dimensions, as described in the section on computer graphic feature analysis, it is straightforward to create an n-dimensional parallelepiped for classification purposes.

We will review how unknown pixels *a* and *b* are assigned to the forest and unclassified categories in Figure 9-16. The computations are summarized in Table 9-8. First, the standard deviation is subtracted and added to the mean of each class and for each band to identify the lower (L_{ck}) and upper (H_{ck}) edges of the parallelepiped. In this case only two bands are used, 4 and 5, resulting in a two-dimensional box. This could be extended to n dimensions or bands. With the lower and upper thresholds for each box identified, it is possible to determine if the brightness value of an input pixel in each band, k, satisfies the criteria of any of the five parallelepipeds. For example, pixel *a* has a value of 40 in both bands 4 and 5. It satisfies the band 4 criteria of class 1 (i.e., $31.27 \leq 40 \leq 41.23$), but it does not satisfy the band 5 criteria. Therefore, the process continues by evaluating the parallelepiped criteria of classes 2 and 3, which are also not satisfied. However, when the brightness values of *a* are compared with class 4 thresholds, we find it satisfies the criteria for band 4 (i.e., $33.99 \leq 40 \leq 44.21$) and band 5 ($29.09 \leq 40 \leq 41.91$). Thus, the pixel is assigned to class 4, forest.

This same logic is applied to classify unknown pixel *b*. Unfortunately, its brightness values of 10 in band 4 and 40 in band 5 never fall within the thresholds of any of the parallelepipeds. Consequently, it is assigned to an

unclassified category. Increasing the size of the thresholds to –2 or 3 standard deviations would increase the size of the parallelepipeds. This might result in point *b* being assigned to one of the classes. However, this same action might also introduce a significant amount of overlap among many of the parallelepipeds resulting in classification error. Perhaps point *b* really belongs to a class that was not trained upon (e.g., dredge spoil).

The parallelepiped algorithm is a computationally efficient method of classifying remote sensor data. Unfortunately, because some parallelepipeds overlap, it is possible that an unknown candidate pixel might satisfy the criteria of more than one class. In such cases it is usually assigned to the first class for which it meets all criteria. A more elegant solution is to take this pixel that can be assigned to more than one class and use a *minimum distance to means* decision rule to assign it to just one class.

Minimum Distance to Means Classification Algorithm

The *minimum distance to means* decision rule is computationally simple and commonly used. When used properly it can result in classification accuracy comparable to other more computationally intensive algorithms such as the maximum likelihood algorithm. Like the parallelepiped algorithm, it requires that the user provide the mean vectors for each class in each band μ_{ck} from the training data. To perform a minimum distance classification, a program must calculate the distance to each mean vector μ_{ck} from each unknown pixel (BV_{ijk}) (Lo and Yeung, 2007). It is possible to calculate this distance using Euclidean distance based on the Pythagorean theorem or "round the block" distance measures (Figure 9-17). In this discussion we demonstrate the method of minimum distance classification using Euclidean distance measurements applied to the two unknown points (*a* and *b*) shown in Figure 9-17.

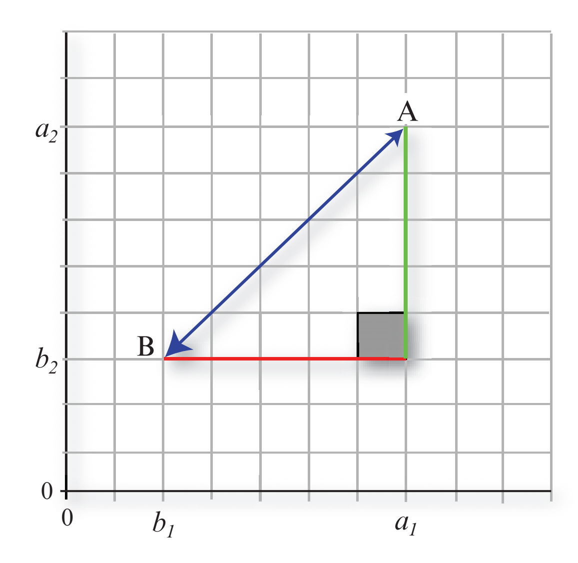
The computation of the Euclidean distance from point *a* (40, 40) to the mean of class 1 (36.7, 55.7) measured in bands 4 and 5 relies on the equation:

$$\text{Dist} = \sqrt{(BV_{ijk} - \mu_{ck})^2 + (BV_{ijl} - \mu_{cl})^2} \qquad (9.15)$$

where μ_{ck} and μ_{cl} represent the mean vectors for class *c* measured in bands *k* and *l*. In our example this would be:

$$\text{Dist}_{a \text{ to class } 1} = \sqrt{(BV_{ij4} - \mu_{1,4})^2 + (BV_{ij5} - \mu_{1,5})^2} \,. \qquad (9.16)$$

The distance from point *a* to the mean of class 2 in these same two bands would be:



Euclidean **distance**

$$D_{AB} = \sqrt{\sum_{i=1}^{2} (a_i - b_i)^2}$$

Round the block **distance**

$$D_{AB} = \sum_{i=1}^{2} |(a_i - b_i)|$$

FIGURE 9–17 The distance used in a *minimum distance to means* classification algorithm can take two forms: the Euclidean distance based on the Pythagorean theorem and the "round the block" distance. The Euclidean distance is more computationally intensive.

$$\text{Dist}_{a \text{ to class } 2} = \sqrt{(BV_{ij4} - \mu_{2,4})^2 + (BV_{ij5} - \mu_{2,5})^2} \,. \qquad (9.17)$$

Notice that the subscript that stands for class *c* is incremented from 1 to 2. By calculating the Euclidean distance from point *a* to the mean of all five classes, it is possible to determine which distance is shortest. Table 9-9 is a list of the mathematics associated with the computation of distances for the five land-cover classes. It reveals that pixel *a* should be assigned to class 4 (forest) because it obtained the minimum distance of 4.59.The same logic can be applied to evaluate the unknown pixel *b*. It is assigned to class 3 (wetland) because it obtained the minimum distance of 15.75. It should be obvious that any unknown pixel will definitely be assigned to one of the five training classes using this algorithm. There will be no unclassified pixels.

Most minimum distance classification algorithms let the analyst specify a distance or threshold from the class means beyond which a pixel will not be assigned to a category even though it is nearest to the mean of that category. For example, if a threshold of 10.0 were specified, point *a* would still be classified as class 4 (forest) because it had a minimum distance of 4.59, which was below the threshold. Conversely, point *b* would not be assigned to class 3 (wetland) because its minimum distance of 15.75 was greater than the 10.0

TABLE 9–9 Example of minimum distance to means classification logic for pixels *a* and *b* in Figure 9-16.

Class	Distance from Pixel *a* (40, 40) to the Mean of Each Class	Distance from Pixel *b* (10, 40) to the Mean of Each Class
1. Residential	$\sqrt{(40-36.7)^2+(40-55.7)^2} = 16.04$	$\sqrt{(10-36.7)^2+(40-55.7)^2} = 30.97$
2. Commercial	$\sqrt{(40-54.8)^2+(40-77.4)^2} = 40.22$	$\sqrt{(10-54.8)^2+(40-77.4)^2} = 58.35$
3. Wetland	$\sqrt{(40-20.2)^2+(40-28.2)^2} = 23.04$	$\sqrt{(10-20.2)^2+(40-28.2)^2} = 15.75$ Assign pixel *b* to this class; it has the minimum distance.
4. Forest	$\sqrt{(40-39.1)^2+(40-35.5)^2} = 4.59$ Assign pixel *a* to this class; it has the minimum distance.	$\sqrt{(10-39.1)^2+(40-35.5)^2} = 29.45$
5. Water	$\sqrt{(40-9.3)^2+(40-5.2)^2} = 46.4$	$\sqrt{(10-9.3)^2+(40-5.2)^2} = 34.8$

threshold. Instead, point *b* would be assigned to an unclassified category.

When more than two bands are evaluated in a classification, it is possible to extend the logic of computing the distance between just two points in *n* space using the equation (Duda et al., 2001):

$$D_{AB} = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} . \quad (9.18)$$

Figure 9-17 demonstrates diagrammatically how this algorithm is used.

Hodgson (1988) identified six additional Euclidean-based minimum distance algorithms that decrease computation time by exploiting two areas: 1) the computation of the distance estimate from the unclassified pixel to each candidate class, and 2) the criteria for eliminating classes from the search process, thus avoiding unnecessary distance computations. Algorithms implementing these improvements were tested using up to 2, 4, and 6 bands of TM data and 5, 20, 50, and 100 classes. All algorithms were more efficient than the traditional Euclidean minimum distance algorithm.

A traditional minimum distance to means classification algorithm was run on the Charleston, SC, Landsat Thematic Mapper dataset using the training data previously described. The results are displayed as a color-coded thematic map in Figure 9-18a. The total number of pixels in each class are summarized in Table 9-10. Error associated with the classification is discussed in Chapter 13.

Nearest-Neighbor Classifiers

The parallelepiped classifier made use of the training class mean and standard deviation statistics for each band. The minimum distance classifier required the mean for each training class for each band. It is also possible to classify an unknown pixel measurement vector into *m* classes using just the training data brightness values in each band (i.e., the mean and standard deviation statistics are not used) and nearest-neighbor distance measurements (Duda et al., 2001; Fuchs et al., 2009). The most common nonparametric nearest-neighbor classifiers are:

- nearest-neighbor,
- *k*-nearest-neighbor, and
- *k*-nearest-neighbor distance-weighted.

The simple *nearest-neighbor classifier* computes the Euclidean distance from the pixel to be classified to the nearest training data pixel in *n*-dimensional feature space and assigns it to that class (Schowengerdt, 2007; Myint et al., 2011).

The nearest-neighbor classifier makes no assumptions about the data being normally distributed. Given a pixel to be classified (*z*) and a training pixel (*t*), the equation to compute the nearest neighbor distance between them is (Jensen et al., 2009):

$$d_{nn} = \sum_{i=1}^{b} (BV_{zi} - BV_{ti})^2 \quad (9.19)$$

where BV_{zi} is the brightness value for unlabeled pixel *z* in band *i*, and BV_{ti} is the brightness value of a training pixel *t* in band *i*. During the pixel assignment, the pixel is given the label of the nearest neighbor that generates

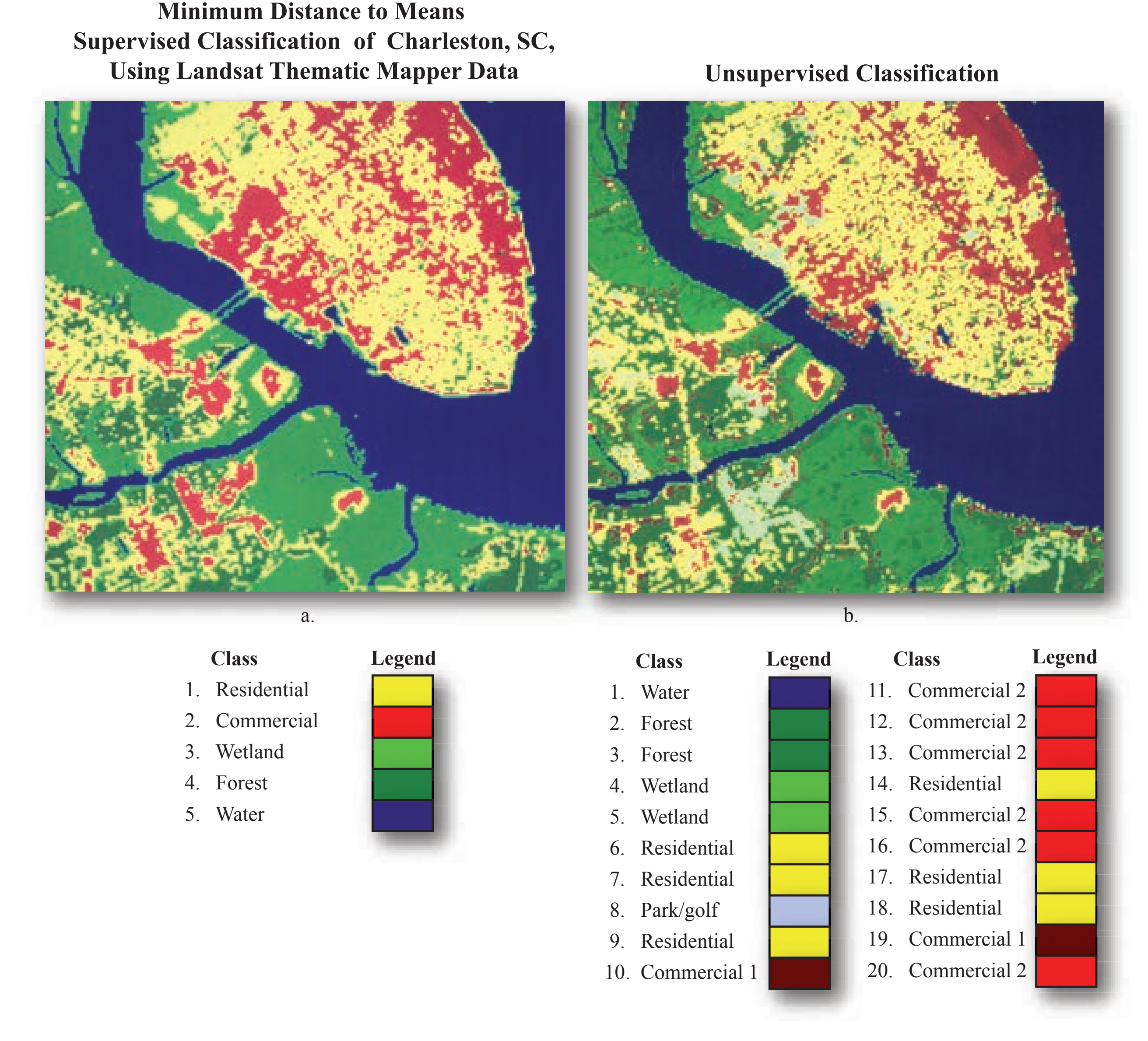


FIGURE 9–18 a) The results of applying a *minimum distance to means* classification algorithm to Landsat Thematic Mapper bands 4 and 5 data of Charleston, SC. Table 9-10 summarizes the number of pixels in each class. b) The results of performing an *unsupervised classification* of the Charleston, SC, Landsat Thematic Mapper imagery using bands 3, 4, and 5. Twenty spectral clusters were extracted and relabeled as information classes according to the criteria shown in Figure 9-29 and Table 9-11. Note that the unsupervised classification extracted a Park/golf class and was able to identify two classes of commercial land cover. Original imagery courtesy of NASA.

the smallest value of d_{nn} among all the training pixels. For example, in the example in Figure 9-19, the nearest-neighbor logic would assign the unknown pixel measurement vector to the commercial class.

The *k-nearest-neighbor classifier* searches away from the pixel to be classified in all directions until it encounters k user-specified training pixels (e.g., $k = 5$). It then assigns the pixel to the class with the majority of pixels encountered. For example, the pixel under investigation in Figure 9-19 would be assigned to the residential class if a k-nearest-neighbor classifier were applied because three of the five training class pixels encountered within the circle are residential.

The *k-nearest-neighbor distance-weighted classifier* uses the same k-nearest neighbors but it weights them according to the distance-weighting logic discussed in Chapter 7. The pixel under investigation is assigned to the training class pixel with the highest total weight. The pixel under investigation in Figure 9-19 would be assigned to the commercial land-use class once again if the k-nearest-neighbor distance-weighted algorithm was applied.

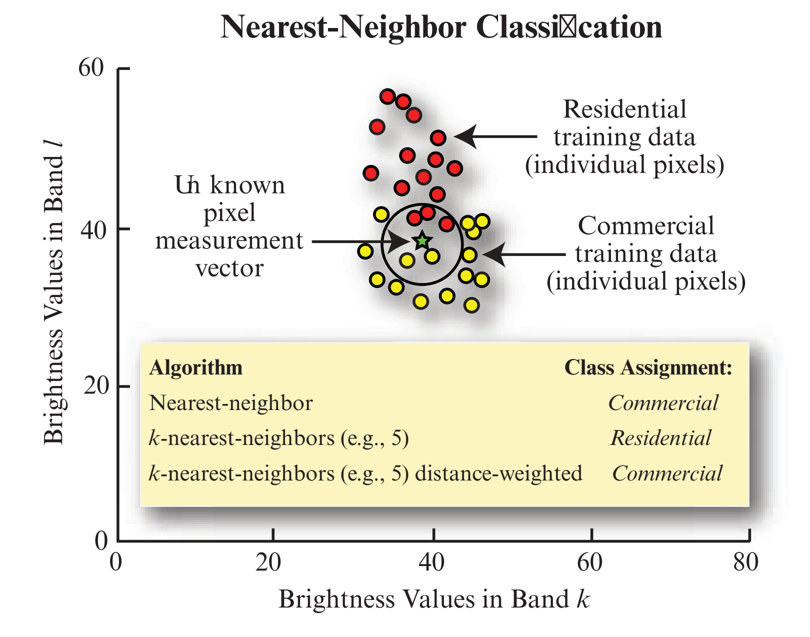


FIGURE 9–19 Hypothetical example of *nearest-neighbor classification.* Simple nearest-neighbor classification locates the nearest training class pixel in *n*-dimensional feature space from the unknown pixel measurement vector under consideration and assigns the unknown pixel the class of the training data. The *k*-nearest-neighbor classifier locates the nearest *k* (e.g., 5) training pixels (no matter what their class) in feature space and assigns the pixel under investigation to the majority class. The *k*-nearest-neighbor distance-weighted algorithm measures the distance to the same *k* training class pixels and weights each of them according to their distance to the pixel under investigation. The pixel is assigned the class with the highest total weight.

TABLE 9–10 Total number of pixels classified into each of the five Charleston, SC, land-cover classes shown in Figure 9-18.

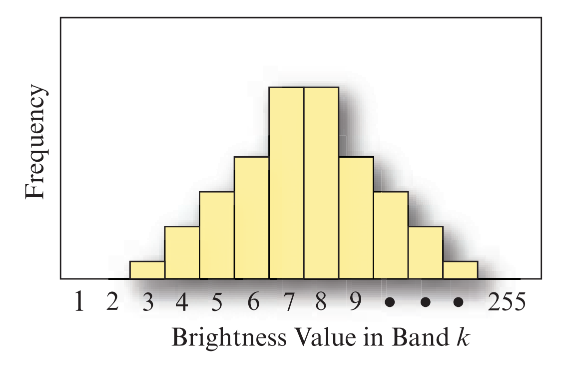
Class	Total Number of Pixels
1. Residential	14,398
2. Commercial	4,088
3. Wetland	10,772
4. Forest	11,673
5. Water	20,509

Nearest-neighbor classifiers can be relatively slow because of the number of distance calculations required between the unknown pixel measurement vector and all of the training pixels in the various bands (Hardin, 1994). But, nearest-neighbor classifiers can yield very useful results if the training data are well separated in *n*-dimensional feature space. Otherwise, it will probably be necessary to use a different algorithm (e.g., minimum distance to means, maximum likelihood, neural network).

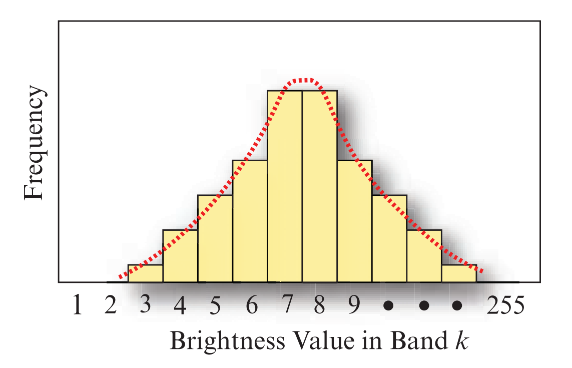
Maximum Likelihood Classification Algorithm

The aforementioned classifiers were based primarily on identifying decision boundaries in feature space based on training class multispectral distance measurements. The **maximum likelihood** decision rule is based on probability. It assigns each pixel having pattern measurements or features *X* to the class *i* whose units are most probable or likely to have given rise to feature vector *X* (Lo and Yeung, 2007; Dalponte et al., 2009). In other words, the probability of a pixel belonging to each of a predefined set of *m* classes is calculated, and the pixel is then assigned to the class for which the probability is the highest. The maximum likelihood decision rule continues to be one of the most widely used supervised classification algorithms (e.g., Campbell and Wynne, 2011; Myint et al., 2011).

The maximum likelihood classification procedure assumes that the training data statistics for each class in each band are normally distributed (i.e., Gaussian). Training data with bi- or *n*-modal histograms in a single band are not ideal. In such cases the individual modes probably represent unique classes that should be trained upon individually and labeled as separate training classes. This should then produce unimodal, Gaussian training class statistics that fulfill the normal distribution requirement.



a. Histogram (data frequency distribution) of forest training data in a single band k.



b. Data distribution approximated by a normal probability density function.

FIGURE 9–20 a) Hypothetical histogram of the forest training data in band k. b) Probability density function computed for the hypothetical training data using Equation 9.20. This function can be used to approximate the frequency of occurrence for every value on the x-axis.

But how do we obtain the probability information we will need from the remote sensing training data we have collected? The answer lies first in the computation of *probability density functions*. We will demonstrate using a single class of training data based on a single band of imagery. For example, consider the hypothetical histogram (data frequency distribution) of forest training data obtained in band k shown in Figure 9-20a. We could choose to store the values contained in this histogram in the computer, but a more elegant solution is to approximate the distribution by a normal probability density function (curve), as shown superimposed on the histogram in Figure 9-20b. The *estimated* probability function for class w_i (e.g., forest) is computed using the equation:

$$p(x|w_i) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma_i} \exp\left[-\frac{1}{2}\frac{(x-\mu_i)^2}{\sigma_i^2}\right] \tag{9.20}$$

where exp [ ] is e (the base of the natural logarithms) raised to the computed power, x is one of the brightness values on the x-axis, μ_i is the *estimated* mean of all the values in the forest training class, and σ_i^2 is the estimated variance of all the measurements in this class. Therefore, we need to store only the mean and variance of each training class (e.g., forest) to compute the probability function associated with any of the individual brightness values in it.

But what if our training data consist of multiple bands of remote sensor data for the classes of interest? Then, we compute an n-dimensional multivariate normal density function using (Swain and Davis, 1978):

$$p(X|w_i) = \frac{1}{(2\pi)^{\frac{n}{2}}|V_i|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(X-M_i)^T V_i^{-1}(X-M_i)\right] \tag{9.21}$$

where $|V_i|$ is the determinant of the covariance matrix, V_i^{-1} is the inverse of the covariance matrix, and $(X - M_i)^T$ is the transpose of the vector $(X - M_i)$. The mean vectors (M_i) and covariance matrix (V_i) for each class are estimated from the training data. For example, consider Figure 9-21 where the bivariate probability density functions of six hypothetical classes are arrayed in red and near-infrared feature space. It is bivariate because two bands are used. Note how the probability density function values appear to be normally distributed (i.e., bell-shaped). The vertical axis is associated with the probability of an unknown pixel measurement vector X being a member of one of the classes. In other words, if an unknown measurement vector has brightness values such that it lies within the wetland region, it has a high probability of being wetland land cover.

If we assume that there are m classes, then $p(X/w_i)$ is the probability density function associated with the unknown measurement vector X, given that X is from a pattern in class w_i (Swain and Davis, 1978). In this case the *maximum likelihood decision rule* becomes (Richards, 2013):

Decide $X \in w_i$ if, and only if,

$$p(X|w_i) \cdot p(w_i) \geq p(X|w_j) \cdot p(w_j) \tag{9.22}$$

for all i and j out of 1, 2,... m possible classes.

Therefore, to classify a pixel in the multispectral remote sensing dataset with an unknown measurement vector X, a maximum likelihood decision rule computes the product $p(X|w_i) \cdot p(w_i)$ for each class and assigns the pattern to the class having the largest product. This assumes that we have some useful information about the prior probabilities of each class i (i.e., $p(w_i)$).

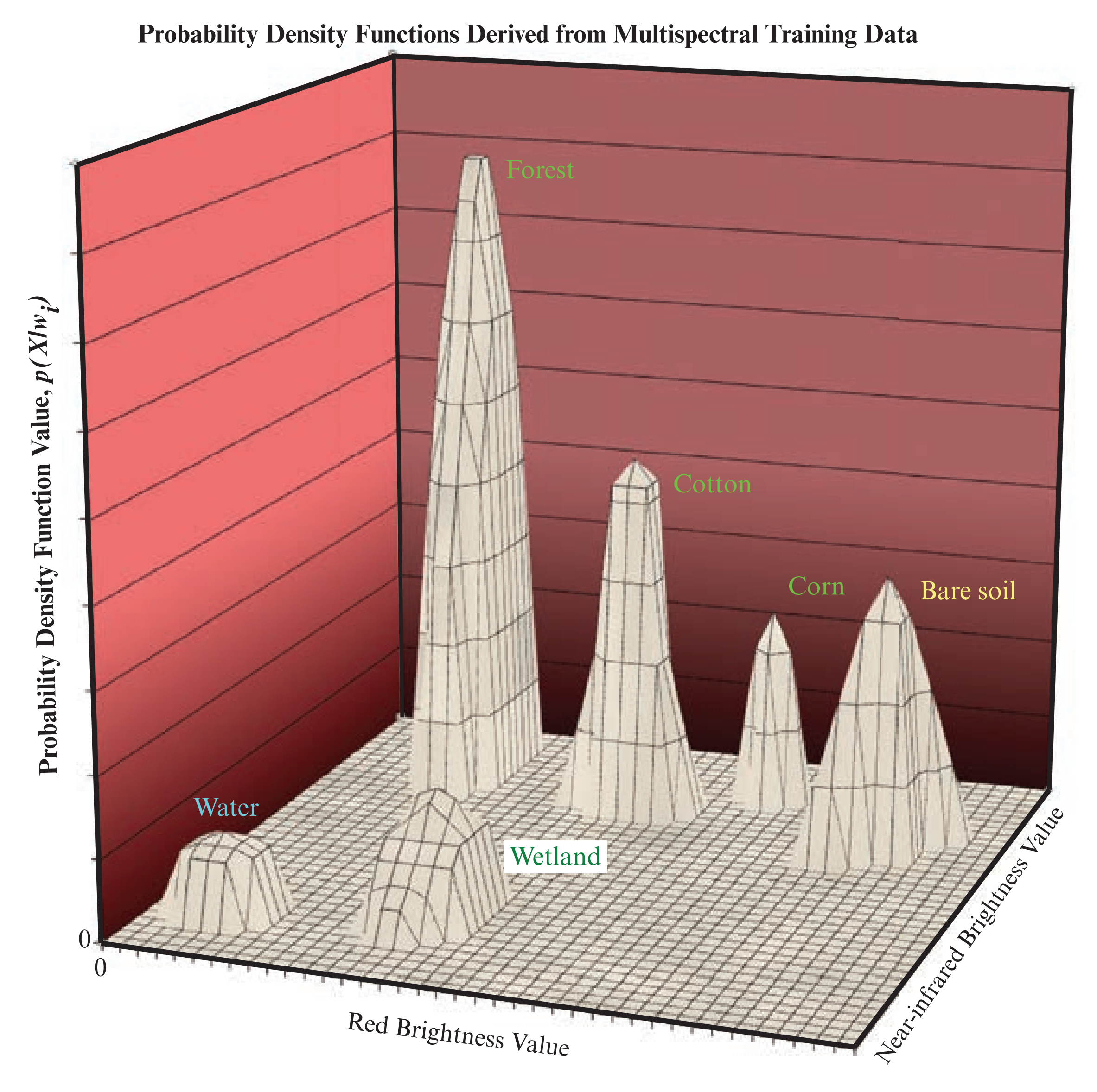


FIGURE 9–21 Example of normally distributed probability density functions associated with six hypothetical land-cover classes using two bands of multispectral data (red and near-infrared). The vertical axis is the probability of an unknown pixel measurement vector X being a member of one of the classes.

Maximum Likelihood Classification Without Prior Probability Information In practice, we rarely have prior information about whether one class (e.g., forest) is expected to occur more frequently in a scene than any other class (e.g., 60% of the scene should be forest). This is called class *a priori* probability information (i.e., $p(w_i)$). Therefore, most applications of the maximum likelihood decision rule assume that each class has an equal probability of occurring in the landscape. This makes it possible to remove the prior probability term [$p(w_i)$] in Equation 9.22 and develop a simplified decision rule that can be applied to the unknown measurement vector X for each pixel in the scene (Dalponte et al., 2009):

Decide unknown measurement vector X is in class i if, and only if,

$$p_i \geq p_j \tag{9.23}$$

for all i and j out of 1, 2, ... m possible classes

and

$$p_i = -\frac{1}{2}\log_e|V_i| - \left[\frac{1}{2}(X - M_i)^T V_i^{-1}(X - M_i)\right] \tag{9.24}$$

where M_i is the mean measurement vector for class i and V_i is the covariance matrix of class i for bands k through l. Therefore, to assign the measurement vector X of an unknown pixel to a class, the maximum likelihood decision rule computes the value p_i for each class. Then it assigns the pixel to the class that has the largest

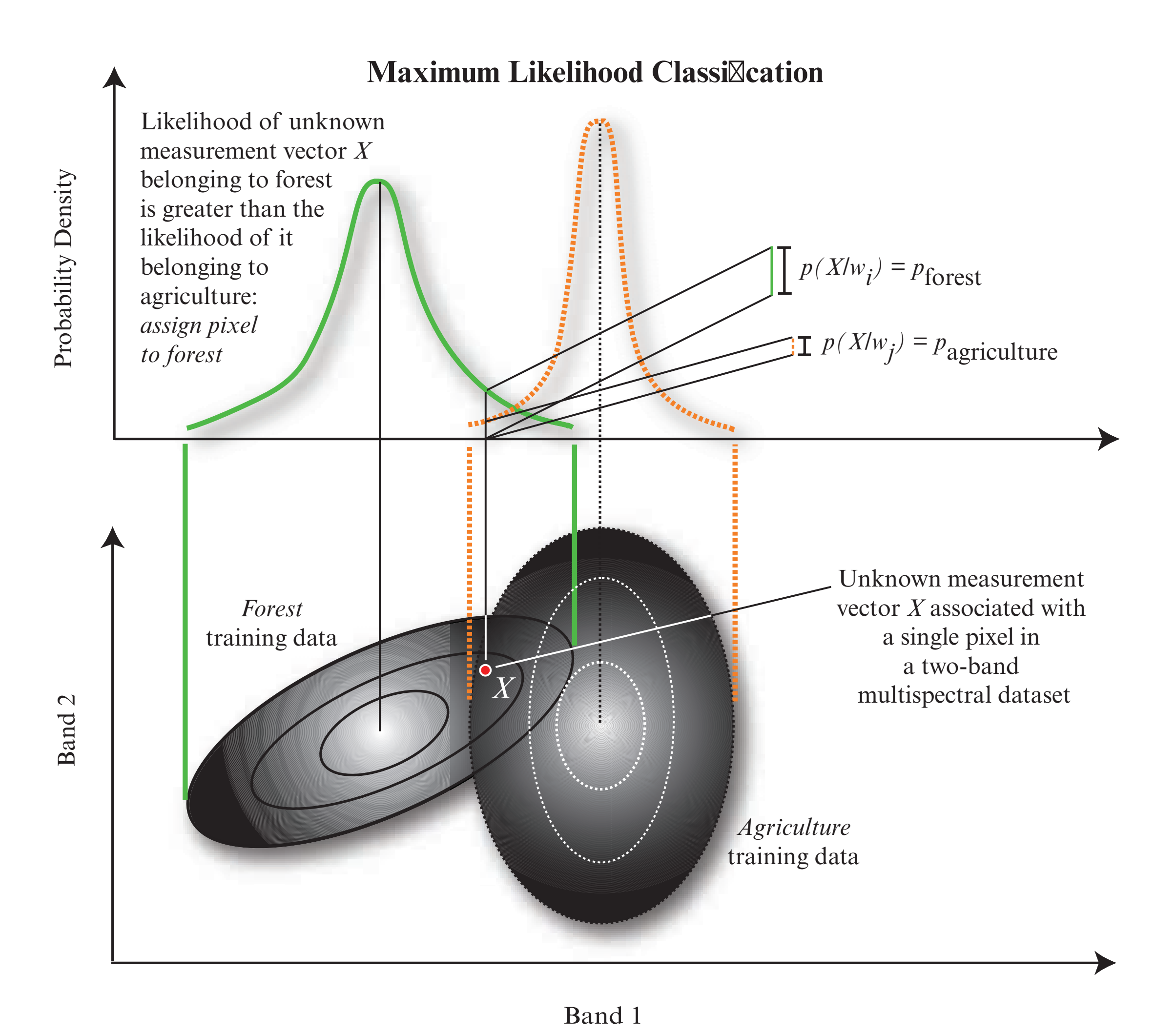


FIGURE 9–22 Hypothetical example of how the maximum likelihood decision rule functions when two training class probability density functions (forest and agriculture) measured in two bands overlap. The unknown measurement vector *X* associated with a single pixel in a two-band dataset would be assigned to forest because the probability density of its measurement vector *X* is greater for forest than for agriculture.

(or maximum) value. This assumes that we have no useful information about the prior probabilities of each of the land-cover classes, that is, every class has an equal probability of occurring throughout the real-world landscape.

Now let us consider the computations required. In the first pass, p_1 is computed with V_1 and M_1 being the covariance matrix and mean vectors for class 1. Next, p_2 is computed with V_2 and M_2 being the covariance matrix and mean vectors for class 2. This continues for all m classes. The pixel or measurement vector X is assigned to the class that produces the largest or maximum p_i. The measurement vector X used in each step of the calculation consists of n elements (the number of bands being analyzed). For example, if six Landsat TM bands (i.e., no thermal band) were being analyzed, each unknown pixel would have a measurement vector X of:

$$X = \begin{bmatrix} BV_{i,j,1} \\ BV_{i,j,2} \\ BV_{i,j,3} \\ BV_{i,j,4} \\ BV_{i,j,5} \\ BV_{i,j,7} \end{bmatrix}. \qquad (9.25)$$

But what happens when the probability density functions of two or more training classes overlap in feature space? For example, consider two hypothetical normally distributed probability density functions associated with forest and agriculture training data measured in bands 1 and 2 (Figure 9-22). In this case, pixel X would be assigned to forest because the probability density of unknown measurement vector X is greater for forest than for agriculture.

Maximum Likelihood Classification with Prior Probability Information Equation 9.23 assumes that each class has an equal probability of occurring in the terrain. Common sense reminds us that in most remote sensing applications there is a high probability of encountering some classes more often than others. For example, in the Charleston scene the probability of encountering various land covers is:

$p(w_1) = p(\text{residential}) = 0.2,$

$p(w_2) = p(\text{commercial}) = 0.1,$

$p(w_3) = p(\text{wetland}) = 0.25,$

$p(w_4) = p(\text{forest}) = 0.1,$ and

$p(w_5) = p(\text{water}) = 0.35.$

Thus, we would expect more pixels to be classified as water simply because it is more prevalent in the image. If we have such information, it is possible to include this valuable prior knowledge in the classification decision. We can do this by weighting each class i by its appropriate *a priori* probability, $p(w_i)$. The equation then becomes:

Decide unknown measurement vector X is in class i if, and only if,

$$p_i \cdot p(w_i) \geq p_j \cdot p(w_j) \tag{9.26}$$

for all i and j out of 1, 2, ... m possible classes

and

$$p_i \cdot p(w_i) = \log_e p(w_i) - \frac{1}{2}\log_e|V_i| - \left[\frac{1}{2}(X - M_i)^T V_i^{-1}(X - M_i)\right].$$

This Bayesian decision rule is identical to the maximum likelihood decision rule except that it does not assume that each class has equal probabilities. *A priori* probabilities have been used successfully as a way of incorporating the effects of relief and other terrain characteristics in improving classification accuracy. The maximum likelihood and Bayes classifications require many more computations per pixel than either the parallelepiped or minimum distance classification algorithms. These algorithms do not always produce superior classification results.



Unsupervised Classification

Unsupervised classification (also commonly referred to as *clustering*) is an effective method of partitioning remote sensor image data in multispectral feature space and extracting land-cover information (Huang, 2002; Exelis ENVI, 2013). Compared to supervised classification, unsupervised classification normally requires only a minimal amount of initial input from the analyst. This is because clustering does not normally require training data.

Unsupervised classification is the process whereby numerical operations are performed that search for natural groupings of the spectral properties of pixels, as examined in multispectral feature space. The clustering process results in a classification map consisting of m spectral classes (Lo and Yeung, 2007). The analyst then attempts *a posteriori* (after the fact) to assign or transform the *spectral* classes into thematic *information* classes of interest (e.g., forest, agriculture, urban). This may not be easy. Some spectral clusters may be meaningless because they represent mixed classes of Earth surface materials. It takes careful thinking by the analyst to unravel such mysteries. The analyst must understand the spectral characteristics of the terrain well enough to be able to label certain clusters as specific information classes.

Hundreds of clustering algorithms have been developed (Duda et al., 2001; Schowengerdt, 2007). Two examples of conceptually simple but not necessarily efficient clustering algorithms will be used to demonstrate the fundamental logic of unsupervised classification of remote sensor data.

Unsupervised Classification Using the Chain Method

The first clustering algorithm that will be discussed operates in a two-pass mode (i.e., it passes through the multispectral dataset two times). In the first pass, the program reads through the dataset and sequentially builds clusters (groups of points in spectral space). A mean vector is then associated with each cluster. In the second pass, a minimum distance to means classification algorithm similar to the one previously described is applied to the whole dataset on a pixel-by-pixel basis whereby each pixel is assigned to one of the mean vectors created in pass 1. The first pass, therefore, automatically creates the cluster signatures to be used by the minimum distance to means classifier.

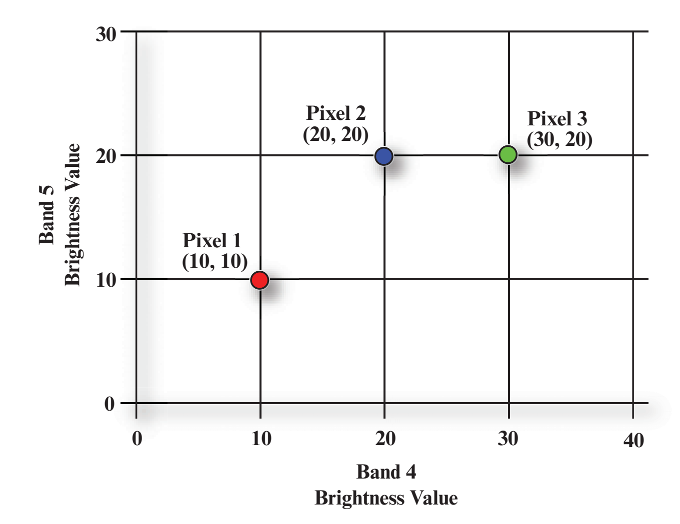


FIGURE 9–23 Original values of pixels 1, 2, and 3 as measured in bands 4 and 5 of the hypothetical remotely sensed data.

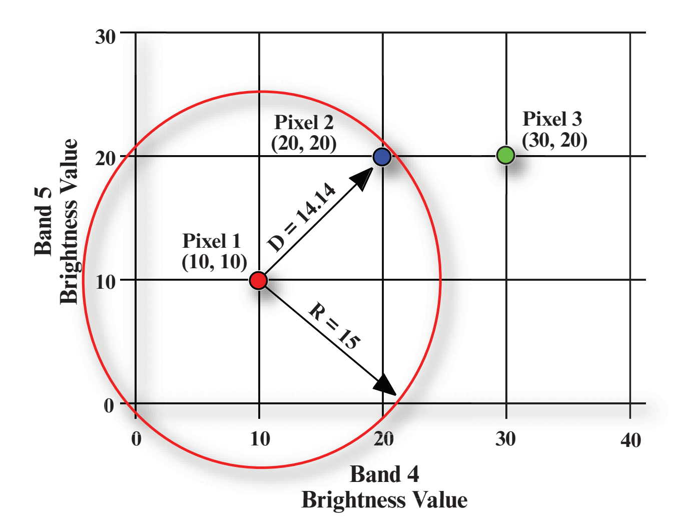


FIGURE 9–24 The distance (*D*) in two-dimensional spectral space between pixel 1 (cluster 1) and pixel 2 (cluster 2) in the first iteration is computed and tested against the value of *R*, the minimum acceptable radius. In this case, *D* does not exceed *R*; therefore, we merge clusters 1 and 2, as shown in Figure 9-25.

Pass 1: Cluster Building

During the first pass, the analyst is required to supply four types of information:

1. *R,* a radius distance in spectral space used to determine when a new cluster should be formed (e.g., when raw remote sensor data are used, it might be set at 15 brightness value units),
2. *C,* a spectral space distance parameter used when merging clusters (e.g., 30 units) when *N* is reached,
3. *N,* the number of pixels to be evaluated between each major merging of the clusters (e.g., 2,000 pixels), and
4. C_{max}, the maximum number of clusters to be identified by the algorithm (e.g., 20 clusters).

These can be set to default values if no initial human interaction is desired.

Starting at the origin of the multispectral dataset (i.e., line 1, column 1), pixels are evaluated sequentially from left to right as if in a chain. After one line is processed, the next line of data is considered. We will analyze the clustering of only the first three pixels in a hypothetical image and label them pixels 1, 2, and 3. The pixels have brightness values in just two bands, 4 and 5. Their spatial relationships in two-dimensional feature space are shown in Figure 9-23.

First, we let the brightness values associated with pixel 1 in the image represent the mean data vector of cluster 1 (i.e., $M_1 = \{10, 10\}$). It is important to remember it is an *n*-dimensional mean data vector with *n* being the number of bands used in the unsupervised classification. In our example, just two bands are being used, so $n = 2$. Because we have not identified all 20 spectral clusters (C_{max}) yet, pixel 2 will be considered as the mean data vector of cluster 2 (i.e., $M_2 = \{20, 20\}$). If the spectral distance *D* between cluster 2 and cluster 1 is greater than *R*, then cluster 2 will remain cluster 2. However, if the spectral distance *D* is less than *R*, then the mean data vector of cluster 1 becomes the average of the first and second pixel brightness values and the weight (or count) of cluster 1 becomes 2 (Figure 9-24). In our example, the distance *D* between cluster 1 (actually pixel 1) and pixel 2 is 14.14. Because the radius *R* was initially set at 15.0, pixel 2 does not satisfy the criteria for being cluster 2 because its distance from cluster 1 is <15. Therefore, the mean data vectors of cluster 1 and pixel 2 are averaged, yielding the new location of cluster 1 at $M_1 = \{15, 15\}$, as shown in Figure 9-25. The spectral distance *D* is computed using the Pythagorean theorem, as discussed previously.

Next, pixel 3 is considered as the mean data vector of cluster 2 (i.e., $M_2 = \{30, 20\}$). The distance from pixel 3 to the revised location of cluster 1, $M_1 = \{15, 15\}$, is 15.81 (Figure 9-25). Because it is >15, the mean data vector of pixel 3 becomes the mean data vector associated with cluster 2.

This cluster accumulation continues until the number of pixels evaluated is greater than *N*. At that point, the program stops evaluating individual pixels and looks closely at the nature of the clusters obtained thus far. It calculates the distance between each cluster and every other cluster. Any two clusters separated by a spectral

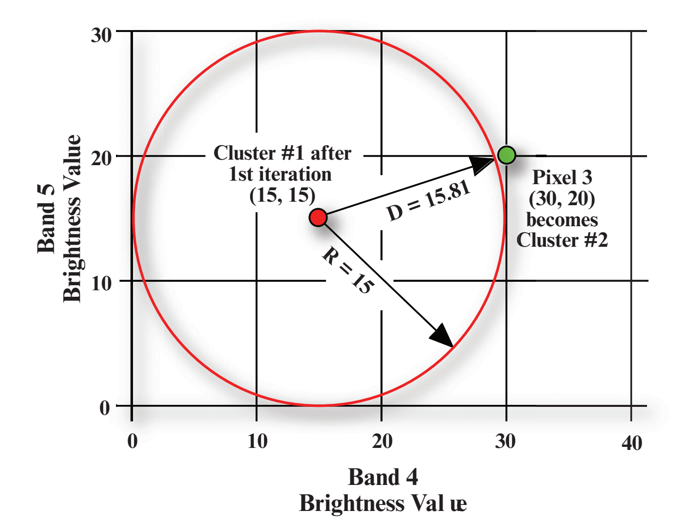


FIGURE 9–25 Pixels 1 and 2 now represent cluster 1. Note that the location of cluster 1 has migrated from (10, 10) to (15, 15), after the first iteration. Now, pixel 3 distance (*D*) is computed to see if it is greater than the minimum threshold, *R*. It is, so pixel location 3 becomes cluster 2. This process continues until all 20 clusters are identified. Then the 20 clusters are evaluated using a distance measure, *C* (not shown in the figure), to merge the clusters that are closest to one another.

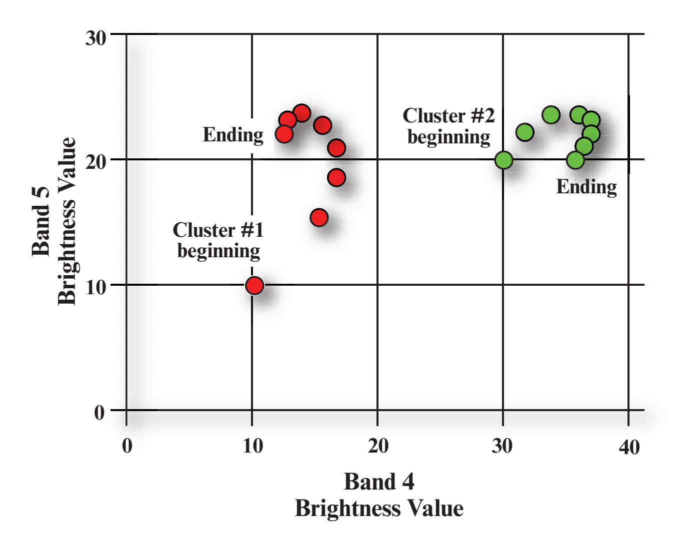


FIGURE 9–26 How clusters migrate during the several iterations of a clustering algorithm. The final ending point represents the mean vector that would be used in phase 2 of the clustering process when the minimum distance classification is performed.

distance less than *C* are merged. Such a new cluster mean vector is the weighted average of the two original clusters, and the weight is the sum of the two individual weights. This proceeds until there are no clusters with a separation distance less than *C*. Then the next pixel is considered. This process continues to iterate until the entire multispectral dataset is examined.

Schowengerdt (2007) suggests that virtually all the commonly used clustering algorithms use iterative calculations to find an optimum set of decision boundaries for the dataset. It should be noted that some clustering algorithms allow the analyst to initially seed the mean vector for several of the important classes. The seed data are usually obtained in a supervised fashion, as discussed previously. Others allow the analyst to use *a priori* information to direct the clustering process.

Some programs do not evaluate every line and every column of the data when computing the mean vectors for the clusters. Instead, they may sample every *i*th row and *j*th column to identify the C_{max} clusters. If computer resources are abundant, then every pixel may be sampled. If resources are scarce, then acceptable results may usually be obtained by sampling the data. Obviously, a great number of computations are performed during this initial pass through the dataset.

A hypothetical diagram showing the cluster migration for our two-band dataset is shown in Figure 9-26. Notice that as more points are added to a cluster, the mean shifts less dramatically since the new computed mean is weighted by the number of pixels currently in a cluster. The ending point is the spectral location of the final mean vector that is used as a signature in the minimum distance classifier applied in pass 2.

Pass 2: Assignment of Pixels to One of the C_{max} Clusters Using Minimum Distance Classification Logic

The final cluster mean data vectors are used in a minimum distance to means classification algorithm to classify all the pixels in the image into one of the C_{max} clusters. The analyst usually produces a cospectral plot display to document where the clusters reside in three-dimensional feature space. It is then necessary to evaluate the location of the clusters in the image, label them if possible, and see if any should be combined. It is usually necessary to combine some clusters. This is where an intimate knowledge of the terrain is critical.

An unsupervised classification of the Charleston, SC, Landsat TM scene is displayed in Figure 9-18b. It was created using TM bands 2, 3, and 4. The analyst stipulated that a total of 20 clusters (C_{max}) be extracted from the data. The mean data vectors for each of the final 20 clusters are summarized in Table 9-11. These mean vectors represented the data used in the minimum distance classification of every pixel in the scene into one of the 20 cluster categories.

TABLE 9–11 Results of clustering on Thematic Mapper Bands 2, 3, and 4 of the Charleston, SC, Landsat TM scene.

Cluster	Percent of Scene	Mean Vector			Class Description	Color Assignment
		Band 2	Band 3	Band 4		
1	24.15	23.14	18.75	9.35	Water	Dark blue
2	7.14	21.89	18.99	44.85	Forest 1	Dark green
3	7.00	22.13	19.72	38.17	Forest 2	Dark green
4	11.61	21.79	19.87	19.46	Wetland 1	Bright green
5	5.83	22.16	20.51	23.90	Wetland 2	Bright green
6	2.18	28.35	28.48	40.67	Residential 1	Bright yellow
7	3.34	36.30	25.58	35.00	Residential 2	Bright yellow
8	2.60	29.44	29.87	49.49	Parks, golf	Gray
9	1.72	32.69	34.70	41.38	Residential 3	Yellow
10	1.85	26.92	26.31	28.18	Commercial 1	Dark red
11	1.27	36.62	39.83	41.76	Commercial 2	Bright red
12	0.53	44.20	49.68	46.28	Commercial 3	Bright red
13	1.03	33.00	34.55	28.21	Commercial 4	Red
14	1.92	30.42	31.36	36.81	Residential 4	Yellow
15	1.00	40.55	44.30	39.99	Commercial 5	Bright red
16	2.13	35.84	38.80	35.09	Commercial 6	Red
17	4.83	25.54	24.14	43.25	Residential 5	Bright yellow
18	1.86	31.03	32.57	32.62	Residential 6	Yellow
19	3.26	22.36	20.22	31.21	Commercial 7	Dark red
20	0.02	34.00	43.00	48.00	Commercial 8	Bright red

Cospectral plots of the mean data vectors for each of the 20 clusters using bands 2 and 3 and bands 3 and 4 are displayed in Figures 9-27 and 9-28, respectively. The 20 clusters lie on a diagonal extending from the origin in the band 2 versus band 3 plot. Unfortunately, the water cluster was located in the same spectral space as forest and wetland when viewed using just bands 2 and 3 (Figure 9-27). Therefore, this scatterplot was not used to *label* or assign the clusters to *information* classes. Conversely, a cospectral plot of bands 3 and 4 mean data vectors was relatively easy to interpret as shown in Figure 9-28.

Cluster labeling is usually performed by interactively displaying all the pixels assigned to an individual cluster on the screen with a color composite of the study area in the background. In this manner it is possible to identify the location and spatial association among clusters. This interactive visual analysis in conjunction with the information provided in the cospectral plot allows the analyst to group the clusters into information classes, as shown in Figure 9-29 and Table 9-11. It is instructive to review some of the logic that resulted in the final unsupervised classification in Figure 9-18b.

Cluster 1 occupied a distinct region of spectral space (Figure 9-29). It was not difficult to assign it to the information class water. Clusters 2 and 3 had high reflectance in the near-infrared (band 4) with low reflectance in the red (band 3) due to chlorophyll absorption. These two clusters were assigned to the forest class and color-coded dark green (refer to Table 9-11). Clusters 4

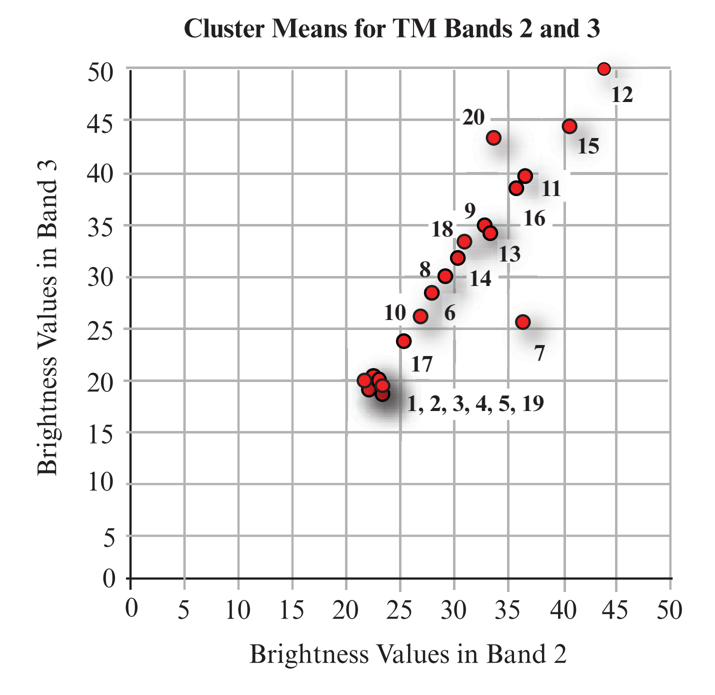


FIGURE 9–27 The mean vectors of the 20 clusters displayed in Figure 9-18b are shown here using only bands 2 and 3. The mean vector values are summarized in Table 9-11. Notice the substantial overlap among clusters 1 through 5 and 19.



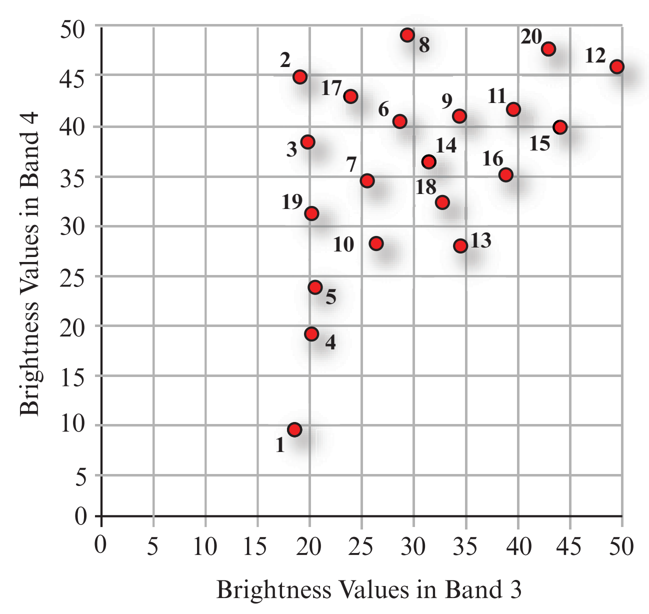


FIGURE 9–28 The mean vectors of the 20 clusters displayed in Figure 9-18b are shown here using only bands 3 and 4 data. The mean vector values are summarized in Table 9-11.

and 5 were situated alone in spectral space between the forest (2 and 3) and water (1) and were comprised of a mixture of moist soil and abundant vegetation. Therefore, it was not difficult to assign both these clusters to a wetland class. They were given different color codes to demonstrate that, indeed, two separate classes of wetland were identified.

Six clusters were associated with residential housing. These clusters were situated between the forest and commercial clusters. This is not unusual since residential housing is composed of a mixture of vegetated and nonvegetated (asphalt and concrete) surfaces, especially at TM spatial resolutions of 30 × 30 meters. Based on where they were located in feature space, the six clusters were collapsed into just two: bright yellow (6, 7, 17) and yellow (9, 14, 18).

Eight clusters were associated with commercial land use. Four of the clusters (11, 12, 15, 20) reflected high amounts of both red and near-infrared energy as commercial land composed of concrete and bare soil often does. Two other clusters (13 and 16) were associated with commercial strip areas, particularly the downtown areas. Finally, there were two clusters (10 and 19) that were definitely commercial but that had a substantial amount of associated vegetation. They were found mainly along major thoroughfares in the residential areas where vegetation is more plentiful. These three subgroups of commercial land use were assigned bright red, red, and dark red, respectively (Table 9-11). Cluster 8 did not fall nicely into any group. It experienced very high near-infrared reflectance and chlorophyll absorption often associated with very well kept lawns or parks. In fact, this is precisely what it was labeled, "parks and golf."

The 20 clusters and their color assignments are shown in Table 9-11 and in Figure 9-18b. There is more information present in this unsupervised classification than in the supervised classification. Except for water, there are at least two classes in each land-use category that could be identified successfully using the unsupervised technique. The supervised classification simply did not sample many of these classes during the initial training process.

Unsupervised Classification Using the ISODATA Method

Another widely used clustering algorithm is the Iterative Self-Organizing Data Analysis Technique (ISODATA). **ISODATA** represents a comprehensive set of heuristic (rule of thumb) procedures that have been incorporated into an iterative classification algorithm (ERDAS, 2013; Pasher and King, 2010; Rich et al., 2010). Many of the steps used in the algorithm are a result of experience gained through experimentation. The ISODATA algorithm is a modification of the *k*-means clustering algorithm, which includes a) merging clusters if their separation distance in multispectral feature

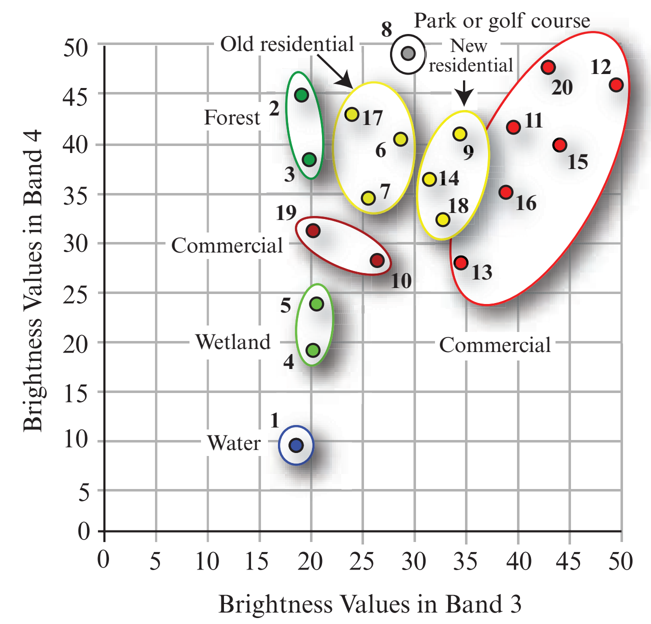


FIGURE 9–29 Grouping (labeling) of the original 20 spectral clusters into information classes used to create the thematic map shown in Figure 9-18b. The labeling was performed by analyzing the mean vector locations in bands 3 and 4.

space is below a user-specified threshold, and b) rules for splitting a single cluster into two clusters (Memarsadeghi et al., 2007; Schowengerdt, 2007; Exelis ENVI, 2013).

ISODATA is self-organizing because it requires relatively little human input. A sophisticated ISODATA algorithm normally requires the analyst to specify the following criteria:

- C_{max}: the maximum number of clusters to be identified by the algorithm (e.g., 20 clusters). However, it is not uncommon for fewer to be found in the final classification map after splitting and merging take place.
- *T:* the maximum percentage of pixels whose class values are allowed to be *unchanged* between iterations. When this number is reached, the ISODATA algorithm terminates. Some datasets may never reach the desired percentage unchanged. If this happens, it is necessary to interrupt processing and edit the parameter.
- *M:* the maximum number of times ISODATA is to classify pixels and recalculate cluster mean vectors. The ISODATA algorithm terminates when this number is reached.
- *Minimum members in a cluster (%):* If a cluster contains less than the minimum percentage of members, it is deleted and the members are assigned to an alternative cluster. This also affects whether a class is going to be split (see maximum standard deviation). The default minimum percentage of members is often set to 0.01.
- *Maximum standard deviation* (σ_{max})*:* When the standard deviation for a cluster exceeds the specified maximum standard deviation and the number of members in the class is greater than twice the specified minimum members in a class, the cluster is split into two clusters. The mean vectors for the two new clusters are the old class centers $\pm 1\sigma$. Maximum standard deviation values between 4.5 and 7 are very typical.
- *Split separation value:* If this value is changed from 0.0, it takes the place of the standard deviation in determining the locations of the new mean vectors plus and minus the split separation value.
- *Minimum distance between cluster means* (*C*)*:* Clusters with a weighted distance less than this value are merged. A default of 3.0 is often used.

Memarsadeghi et al. (2007) summarize the characteristics of the fundamental ISODATA clustering algorithm and then provide a fast implementation. They provide the C^{++} code for the fundamental algorithm as well as the fast implementation at http://www.cs.umd.edu/~mount/Projects/ISODATA/.

ISODATA Initial Arbitrary Cluster Allocation

ISODATA is iterative because it makes a large number of passes through the remote sensing dataset until specified results are obtained, instead of just two passes. Also, ISODATA does not allocate its initial mean vectors based on the analysis of pixels in the first line of data the way the two-pass algorithm does. Rather, an initial arbitrary assignment of all C_{max} clusters takes place along an *n*-dimensional vector that runs between very specific points in feature space. The region in feature space is defined using the mean, μ_k, and standard deviation, σ_k, of each band in the analysis. A hypothetical two-dimensional example using bands 3 and 4 is presented in Figure 9-30a, in which five mean vectors are distributed along the vector beginning at location $\mu_3 - \sigma_3, \mu_4 - \sigma_4$ and ending at $\mu_3 + \sigma_3, \mu_4 + \sigma_4$. This method of automatically seeding the original C_{max} vectors makes sure that the first few lines of data do not bias the creation of clusters. Note that the two-dimensional parallelepiped (box) does not capture all the possible band 3 and 4 brightness value combinations present in the scene. The location of the initial C_{max} mean vectors (Figure 9-30a) should move about somewhat to partition the feature space better. This takes place in the first and subsequent iterations. Huang (2002) developed an efficient version of ISODATA that provides improved automatic location of the initial mean vectors (i.e., seed clusters).

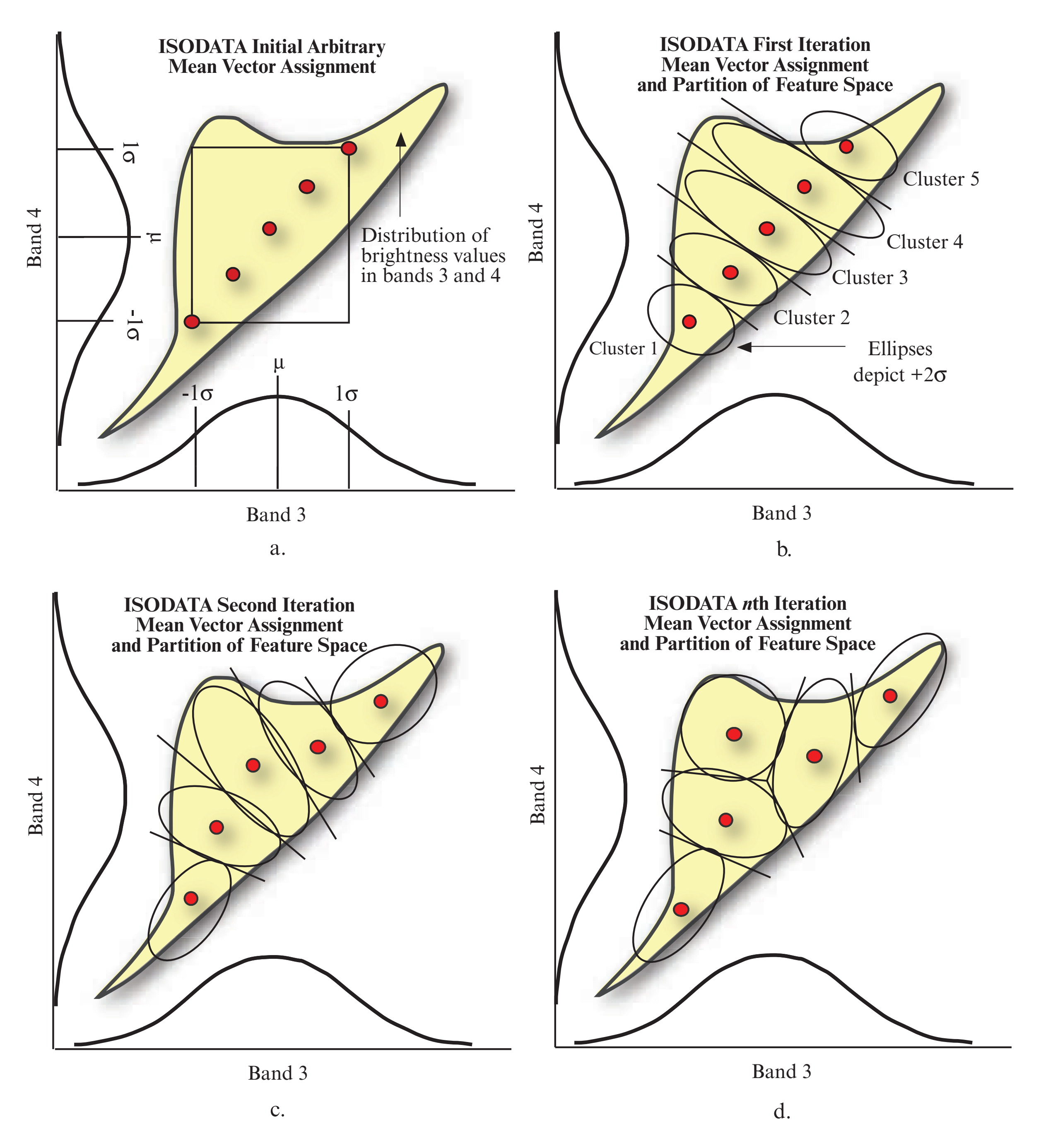


FIGURE 9–30 a) ISODATA initial distribution of five hypothetical mean vectors using -1σ standard deviations in both bands as beginning and ending points. b) In the first iteration, each candidate pixel is compared to each cluster mean and assigned to the cluster whose mean is closest in Euclidean distance. c) During the second iteration, a new mean is calculated for each cluster based on the actual spectral locations of the pixels assigned to each cluster, instead of the initial arbitrary calculation. This involves analysis of several parameters to merge or split clusters. After the new cluster mean vectors are selected, every pixel in the scene is assigned to one of the new clusters. d) This split–merge–assign process continues until there is only a small amount of change in class assignment between iterations (the T threshold is reached) or the maximum number of iterations is reached (M).

ISODATA First Iteration

With the initial C_{max} mean vectors in place, a pass is made through the database beginning in the upper left corner of the matrix. Each candidate pixel is compared to each cluster mean and assigned to the cluster whose mean is closest in Euclidean distance (Figure 9-30b). This pass creates an actual classification map consisting of C_{max} classes. It should be noted that some image processing systems process data line by line, and others process the data in a block or tiled data structure. The way that ISODATA is instructed to process the data

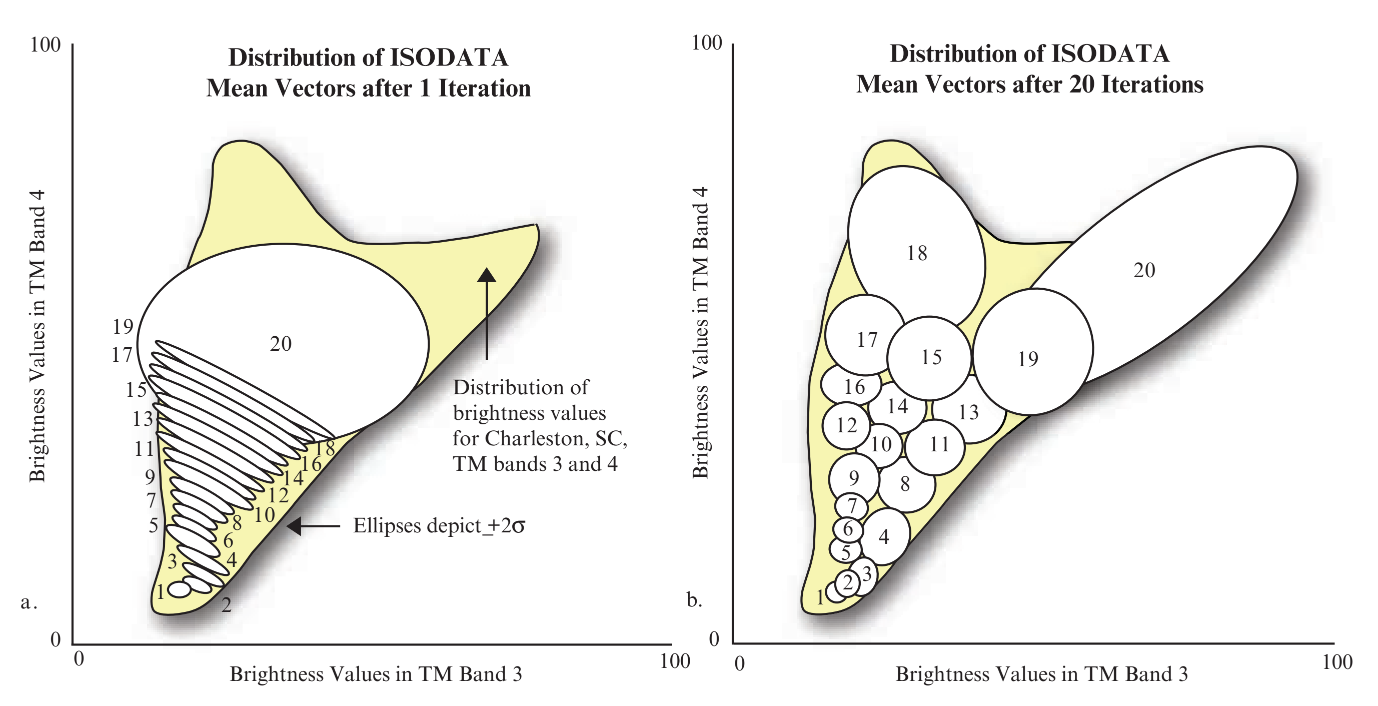


FIGURE 9–31 a) Distribution of 20 ISODATA mean vectors after just one iteration using Landsat TM band 3 and 4 data of Charleston, SC. Notice that the initial mean vectors are distributed along a diagonal in two-dimensional feature space according to the -2σ standard deviation logic discussed. b) Distribution of 20 ISODATA mean vectors after 20 iterations. The bulk of the important feature space (the yellow background) is partitioned rather well after just 20 iterations.

(e.g., line by line or block by block) will have an impact on the creation of the mean vectors.

ISODATA Second to *M*th Iteration

After the first iteration, a new mean for each cluster is calculated based on the actual spectral locations of the pixels assigned to each cluster, instead of on the initial arbitrary calculation. This involves analysis of the following parameters: minimum members in a cluster (%), maximum standard deviation (σ_{max}), split separation, and minimum distance between cluster means (C). Then the entire process is repeated with each candidate pixel once again compared to the new cluster means and assigned to the nearest cluster mean (Figure 9-30c). Sometimes individual pixels do not change cluster assignment. This iterative process continues (Figure 9-30d) until there is 1) little change in class assignment between iterations (i.e., the T threshold is reached), or 2) the maximum number of iterations is reached (M). The final file is a matrix with C_{max} clusters in it, which must be labeled and recoded to become useful land-cover information. The fact that the initial mean vectors are situated throughout the heart of the existing data is superior to initiating clusters based on finding them in the first line of the data.

The iterative ISODATA algorithm is relatively slow, and image analysts are notoriously impatient. Analysts must allow the ISODATA algorithm to iterate enough times to generate meaningful mean vectors.

ISODATA Example 1 An ISODATA classification was performed using the Charleston, SC, Landsat TM band 3 and 4 data. The locations of the clusters (mean -2σ) after one iteration are shown in Figure 9-31a. The clusters are superimposed on the distribution of all brightness values found in TM bands 3 and 4. The location of the final mean vectors after 20 iterations is shown in Figure 9-31b. The ISODATA algorithm has partitioned the feature space effectively. Requesting more clusters (e.g., 100) and allowing more iterations (e.g., 500) would partition the feature space even better. A classification map example is not provided because it would not be dramatically different from the results of the two-pass clustering algorithm since so few clusters were requested.

ISODATA Example 2 An ISODATA algorithm was applied to two bands of HyMap hyperspectral data (red and near-infrared) of an area near North Inlet, SC. The goal was to extract six land-cover classes from the data: water, wetland, roof/asphalt, forest, bare soil, and golf fairway. Only 10 clusters were extracted using ISODATA for demonstration purposes. The locations of the 10 clusters (actually mean vectors) after 1–5 and 10 iterations are displayed in Figure 9-32. The red and near-infrared spectral characteristics of the final 10

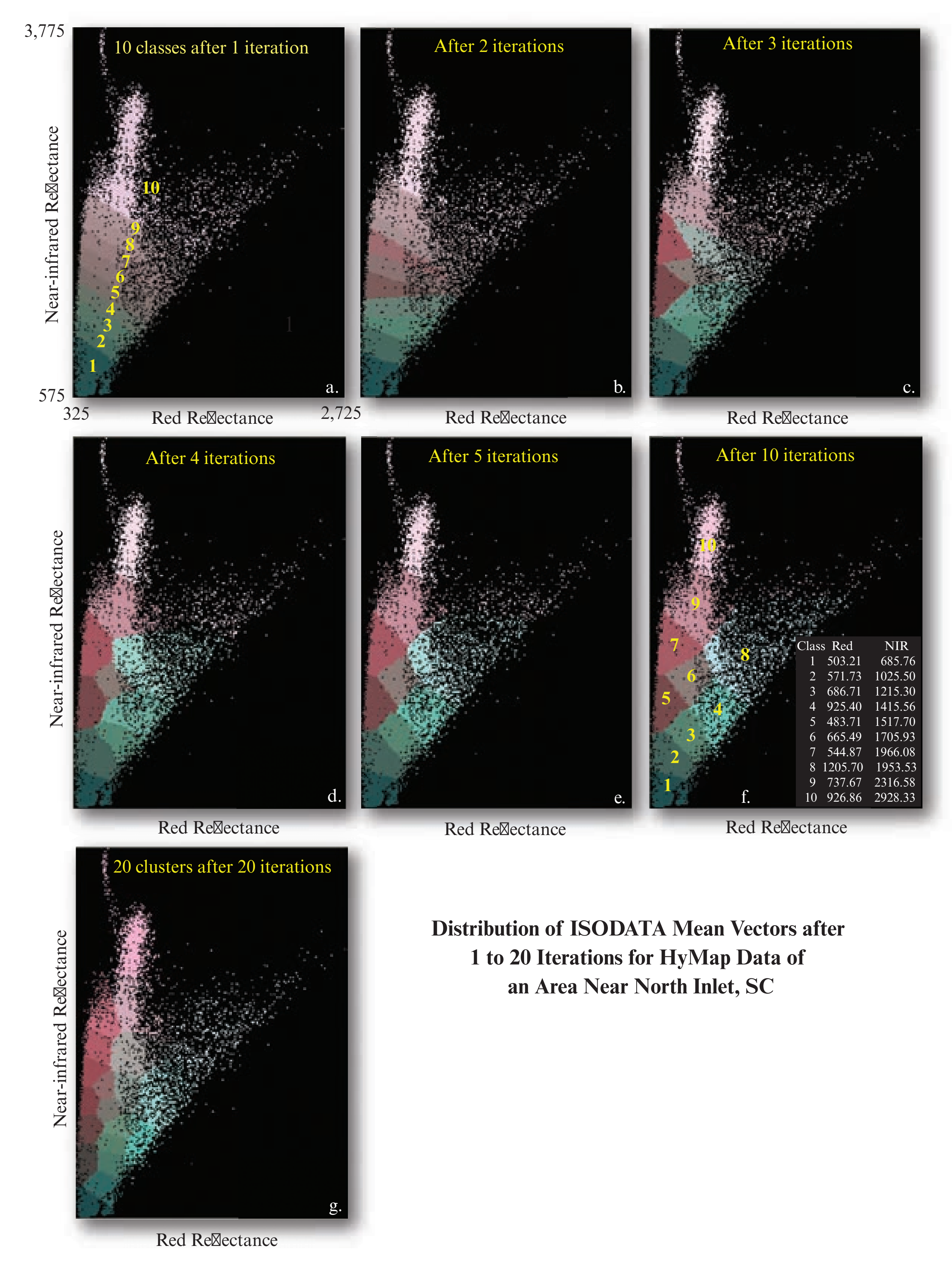
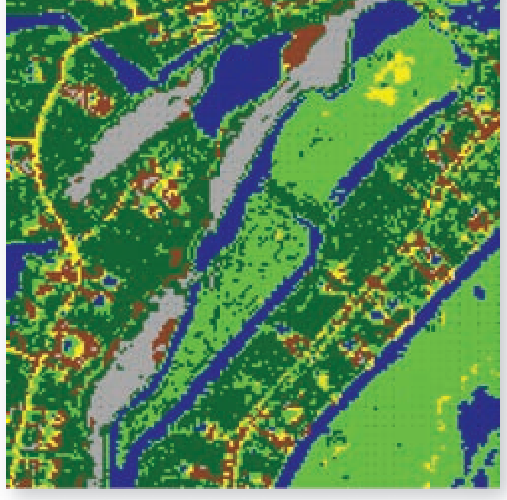


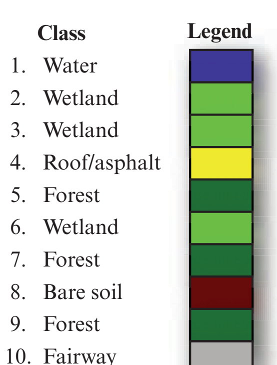
FIGURE 9–32 a–f) Distribution of 10 ISODATA mean vectors after 1–5 and 10 iterations using just two bands of HyMap hyperspectral data of an area near North Inlet, SC. The 10 mean vectors after 10 iterations were used to produce the thematic map shown in Figure 9-33. g) Typically, more than 20 clusters are requested when using ISODATA. This graphic shows the distribution of the two-dimensional feature space for 20 clusters after 20 iterations.

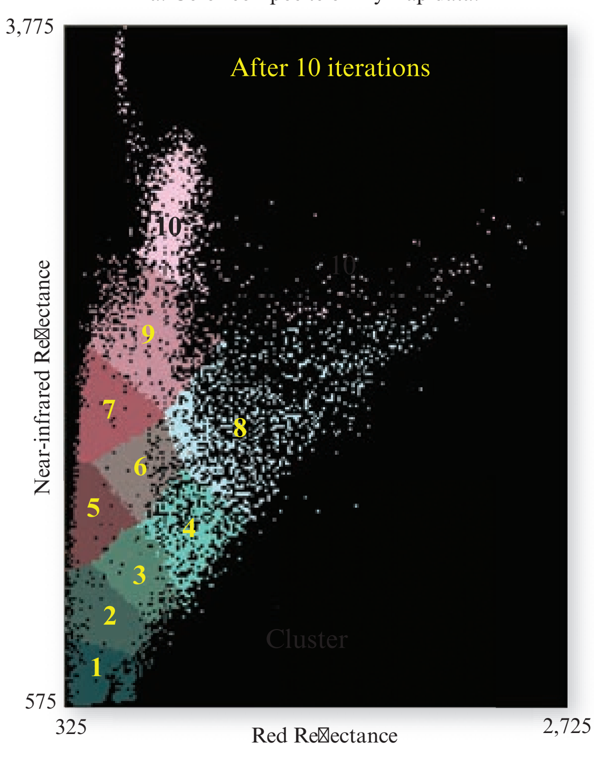
Information Classes Derived from an ISODATA Unsupervised Classification Using 10 Iterations and 10 Mean Vectors of an Area Near North Inlet, SC


a. Color composite of HyMap data.




c. Classification map derived from 10 ISODATA clusters.





b. Final location of 10 ISODATA mean vectors after 10 iterations.

FIGURE 9–33 ISODATA classification of HyMap remote sensor data of an area near North Inlet, SC. a) Color composite image. b) Location of 10 mean vectors in red and near-infrared feature space after 10 iterations. c) Classification map derived by labeling the 10 spectral classes as information classes.

mean vectors after 10 iterations are found in Figure 9-32f. Typically, 20 or more clusters are requested. Therefore, Figure 9-32g shows where 20 clusters reside after 20 iterations.

The 10 mean vectors shown in Figure 9-32f were used in the final pass of the ISODATA algorithm to create a thematic map consisting of 10 spectral classes. These spectral classes were evaluated and labeled as the 10 information classes shown in Figure 9-33c. The final step in the process would be to recode the 10 information classes into just 6 classes to create a normal 6-class legend and to compute the hectares in each land-cover class.

Unsupervised Cluster Busting

It is common when performing unsupervised classification using the chain algorithm or ISODATA to generate *n* clusters (e.g., 100) and have no confidence in labeling *q* of them to an appropriate information class (let us say 30 in this example). This is because 1) the terrain within the IFOV of the sensor system contained at least two types of terrain, causing the pixel to exhibit spectral characteristics unlike either of the two terrain components, or 2) the distribution of the mean vectors generated during the unsupervised classification process was not good enough to partition certain important portions of feature space. When this occurs, it may be possible to perform *cluster busting* if in fact there is still some unextracted information of value in the remote sensing dataset.

First, all the pixels associated with the *q* clusters (30 in a hypothetical example) that are difficult to label (e.g., mixed clusters 13, 22, 45, 92, etc.) are all recoded to a value of 1 and a binary mask file is created. A mask program is then run using 1) the binary mask file, and 2) the original remote sensor data file. The output of the mask program is a new multiband image file consisting of only the pixels that could not be adequately labeled during the initial unsupervised classification. The analyst then performs a new unsupervised classification on this file, perhaps requesting an additional 25 clusters. The analyst displays these clusters using standard techniques and keeps as many of these new clusters as possible (e.g., 15). Usually, there are still some clusters that contain mixed pixels, but the proportion definitely goes down. The analyst may want to iterate the process one more time to see if an additional unsupervised classification breaks out additional clusters. Perhaps five good clusters are extracted during the final iteration.

In this hypothetical example, the final cluster map would be composed of the 70 good clusters from the initial classification, 15 good clusters from the first cluster-busting pass (recoded as values 71 to 85), and 5 from the second cluster-busting pass (recoded as values 86 to 90). The final cluster map file may be put together using a simple GIS maximum dominate function. The final cluster map is then recoded to create the final classification map.



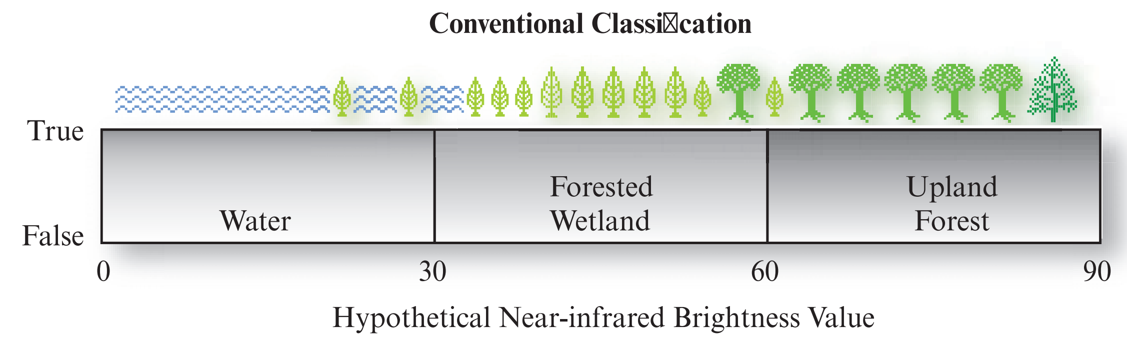
Fuzzy Classification

Geographical information (including remotely sensed data) is imprecise, meaning that the boundaries between different phenomena are fuzzy, or there is heterogeneity within a class, perhaps due to differences in species, health, age, and so forth.

For example, terrain in the southeastern United States often exhibits a gradual transition from water to forested wetland to deciduous upland forest as shown in Figure 9-34a. Normally, the greater the canopy closure, the greater the amount of near-infrared energy reflected from within the IFOV of a pixel along this continuum. Also, the greater the proportion of water in a pixel, the more near-infrared radiant flux absorbed. A hard classification algorithm applied to these remotely sensed data collected along this continuum would be based on classical set theory, which requires precisely defined set boundaries for which an element (e.g., a pixel) is either a member (true = 1) or not a member (false = 0) of a given set. For example, if we made a classification map using just a single near-infrared band (i.e., one-dimensional density slicing), the decision rules might be as shown in Figure 9-34a: 0 to 30 = water, 31 to 60 = forested wetland, and 61 to 90 = upland forest. The classic approach creates three discrete classes with specific class ranges, and no intermediate situations are allowed. Thus, using classical set theory, an unknown measurement vector may be assigned to one and only one class (Figure 9-34a). But everyone knows that the phenomena grade into one another and that mixed pixels are present, especially around the values of 24 to 36 and 55 to 70, as shown in the figure. Clearly, there needs to be a way to make the classification algorithms more sensitive to the imprecise (fuzzy) nature of the real world.

Fuzzy set theory provides some useful tools for working with imprecise data (Wang, 1990a,b; Lu and Weng, 2007; Phillips et al., 2011). Fuzzy set theory is better suited for dealing with real-world problems than traditional logic because most human reasoning is imprecise and is based on the following logic. First, let X be a universe whose elements are denoted x. That is, $X = \{x\}$. As previously mentioned, membership in a classical set A of X is often viewed as a binary characteristic function x_A from X {0 or 1} such that $x_A(x) = 1$ if and only if $x \in A$. Conversely, a *fuzzy set B* in X is characterized by a *membership function* f_B that associates with each x a real number from 0 to 1. The closer the value of $f_B(x)$ is to 1, the more x belongs to B. Thus, a fuzzy set does not have sharply defined boundaries, and a set element (a pixel in our case) may have partial membership in several classes (Campbell and Wynne, 2011; Dronova et al., 2011).

So how is fuzzy logic used to perform image classification? Figure 9-34b illustrates the use of fuzzy classification logic to discriminate among the three hypothetical land covers. The vertical boundary for water at brightness value 30 (Figure 9-34a) is replaced by a graded





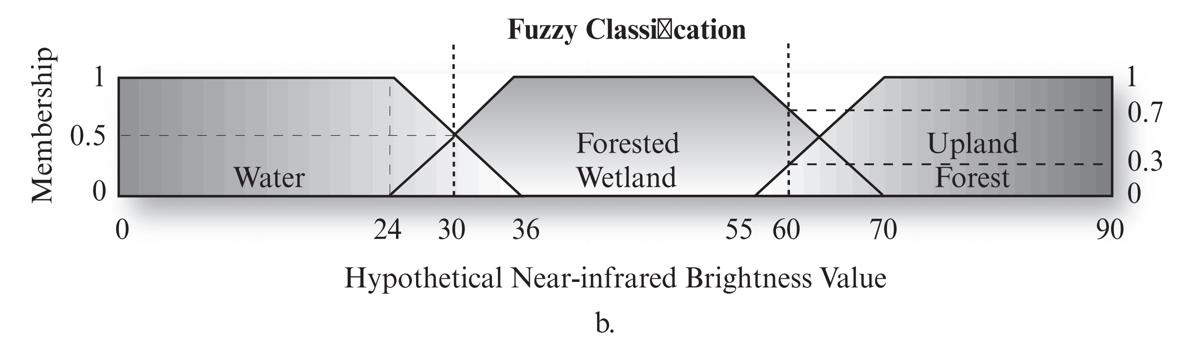


FIGURE 9–34 a) Conventional hard classification rules applied to discriminate among three land-cover classes. The terrain icons suggest that there is a gradual transition in near-infrared brightness value as one progresses from water to forested wetland to upland forest. A remote sensing system would be expected to record radiant flux from mixed pixels at the interface between the major land-cover types. Mixed pixels may also be encountered within a land-cover type as a result of differences in species, age, or functional health of vegetation. Despite these fuzzy conditions, a hard classification would simply assign a pixel to one and only one class. b) The logic of a fuzzy classification. In this hypothetical example, a pixel having a near-infrared brightness value of <24 would have a *membership grade* value of 1.0 in water and 0 in both forested wetland and upland forest. Similarly, a brightness value of 60 would have a graded value of 0.70 for forested wetland, 0.30 for upland forest, and 0 for water. The membership grade values provide information on mixed pixels and may be used to classify the image using various types of logic.

boundary that represents a gradual transition from water to forested wetland (Figure 9-34b). In the language of fuzzy set theory, *BV*s of less than 24 have a *membership grade* of 1 for water, and those greater than about 70 have a membership grade of 1 for upland forest. At several other locations a *BV* may have a membership grade in two classes. For example, at *BV* 30 we have membership grades of 0.5 water and 0.5 of forested wetland. At *BV* 60 the membership grades are 0.7 for forested wetland and 0.3 for upland forest. This membership grade information may be used by the analyst to create a variety of classification maps.

Fuzzy set theory is not a panacea, but it does offer significant potential for extracting information on the makeup of the biophysical materials within a mixed pixel, a problem that will always be with us. Fortunately, **Spectral Mixture Analysis** (SMA) can be used to extract information from mixed pixels (e.g., Song, 2005; Lu and Weng, 2007). It evaluates each pixel spectrum as a linear combination of a set of *endmember* spectra (i.e., pure homogeneous materials such as water, asphalt, closed vegetation canopy, bare soil) (Pu et al., 2008; Colditz et al., 2011). The output of SMA is typically presented in the form of fraction images, with one image for each endmember spectrum, representing the area proportions of the endmembers within the pixel. Thus, SMA provides fuzzy information. SMA is one of the most important hyperspectral image analysis functions and is described in detail in Chapter 11.



Object-Based Image Analysis (OBIA) Classification

The twenty-first century has witnessed the development of remote sensing systems such as IKONOS, QuickBird, GeoEye-1, WorldView-2 and others that produce data with spatial resolutions of $\leq 1 \times 1$ m. Unfortunately, classification algorithms based on single-pixel analysis often are not capable of extracting the information desired from high spatial resolution remote sensor data (Pena-Barragan et al., 2011; Textron Systems, 2014). For example, the spectral complexity of urban land-cover materials can result in limitations

when using per-pixel analysis for the classification of human-made materials such as roads and roofs and natural materials such as vegetation, soil, and water (Herold et al., 2002, 2003; Myint et al., 2011). Furthermore, a significant but usually ignored problem with per-pixel characterization of land cover is that a substantial proportion of the signal apparently coming from the land area represented by a pixel comes from the surrounding terrain (Townshend et al., 2000). Improved algorithms are needed that take into account not only the spectral characteristics of a single pixel but those of the surrounding (contextual) pixels. In addition, we need information about the spatial characteristics of the surrounding pixels so that we can identify areas (or segments) of pixels that are homogeneous (Frohn and Hao, 2006) and those that have changed (e.g., Im et al., 2008).

Geographic Object-Based Image Analysis and Classification

This need has given rise to the creation of **geographic object-based image analysis (GEOBIA).** The algorithms typically incorporate both spectral and spatial information in the image segmentation phase. The result is the creation of *image objects* defined as individual areas with shape and spectral homogeneity (Benz, 2001; Blaschke, 2010; Blaschke et al., 2014), which one may recognize as segments or patches in the landscape ecology literature (Frohn and Hao, 2006; Im et al., 2008). In many instances, extracted image objects can provide a greater number of meaningful features for image classification (Wang et al., 2010; Liu and Xia, 2010; Textron Systems, 2014). The objects can also be developed from any spatially distributed variable (e.g., elevation, slope, aspect, population density). Homogeneous image objects are then analyzed using traditional classification algorithms (e.g., minimum distance, maximum likelihood), knowledge-based approaches and/or fuzzy classification logic.

There are many algorithms that can be used to segment an image into relatively homogeneous image objects. Most can be grouped into two classes: edge-based algorithms and area-based algorithms. Unfortunately, the majority do not incorporate both spectral and spatial information, and very few have been used for remote sensing digital image classification.

A very important approach to remote sensing image segmentation was developed by Baatz and Schape (2000). The image segmentation involves looking at individual pixel values and their neighbors to compute a (Baatz et al., 2001):

- color criterion (h_{color}), and
- a shape or spatial criterion (h_{shape}).

These two criteria are then used to create image objects (patches) of relatively homogeneous pixels in the remote sensing dataset using the general segmentation function (S_f) (Baatz et al., 2001; Definiens, 2003, 2007):

$$S_f = w_{color} \cdot h_{color} + (1 - w_{color}) \cdot h_{shape} \qquad (9.27)$$

where the user-defined weight for spectral color versus shape is $0 \le w_{color} \le 1$. If the user wants to place greater emphasis on the spectral (color) characteristics in the creation of homogeneous objects (patches) in the dataset, then w_{color} is weighted more heavily (e.g., w_{color} = 0.8). Conversely, if the spatial characteristics of the dataset are believed to be more important in the creation of the homogeneous patches, then shape should be weighted more heavily.

Spectral (i.e., color) heterogeneity (h) of an *image object* is computed as the sum of the standard deviations of spectral values of each layer (σ_k) (i.e., band) multiplied by the weights for each layer (w_k) (Kuehn et al., 2002; Definiens, 2003):

$$h = \sum_{k=1}^{m} w_k \cdot \sigma_k . \qquad (9.28)$$

The *col orcriterion* is computed as the weighted mean of all changes in standard deviation for each channel k of the m band remote sensing dataset. The standard deviation σ_k are weighted by the object sizes n_{ob} (Definiens, 2003):

$$h = \sum_{k=1}^{m} w_k [n_{mg} \cdot \sigma_k^{mg} - (n_{ob1} \cdot \sigma_k^{ob1} + n_{ob2} \cdot \sigma_k^{ob2})] \qquad (9.29)$$

where mg means merge.

The *shape criterion* is computed using two landscape ecology metrics: compactness and smoothness. Heterogeneity as deviation from a *compact* shape (cpt) is described by the ratio of the pixel perimeter length l and the square root of the number of pixels n forming an image object (i.e., a patch):

$$\text{cpt} = \frac{l}{\sqrt{n}} . \qquad (9.30)$$

Shape heterogeneity may also be described as *smoothness,* which is the ratio of the pixel perimeter length l and the shortest possible border length b of a box bounding the image object (i.e., a patch) parallel to the raster:

$$\text{smooth} = \frac{l}{b} . \qquad (9.31)$$

The shape criterion incorporates these two measurements using the equation (Definiens, 2003, 2007):

$$h_{shape} = w_{cpt} \cdot h_{cpt} + (1 - w_{cpt}) \cdot h_{smooth} \tag{9.32}$$

where $0 \leq w_{cpt} \leq 1$ is the user-defined weight for the compactness criterion. The change in shape heterogeneity caused by each merge is evaluated by calculating the difference between the situation after and before image objects (*ob*) are merged. This results in the following algorithms for computing roughness and smoothness (Definiens, 2003):

$$h_{cpt} = n_{mg} \cdot \frac{l_{mg}}{\sqrt{n_{mg}}} - \left(n_{ob1} \cdot \frac{l_{ob1}}{\sqrt{n_{ob1}}} + n_{ob2} \cdot \frac{l_{ob2}}{\sqrt{n_{ob2}}} \right) \tag{9.33}$$

$$h_{smooth} = n_{mg} \cdot \frac{l_{mg}}{b_{mg}} - \left(n_{ob1} \cdot \frac{l_{ob1}}{b_{ob1}} + n_{ob2} \cdot \frac{l_{ob2}}{b_{ob2}} \right) \tag{9.34}$$

where n is the object size in pixels.

A *pixel neighborhood* function may be used to determine whether an image object should be grown or whether a new image object should be created (Definiens, 2003). The logic associated with plane 4 and diagonal 8 neighborhood functions is shown in Figure 9-35. In this example, the plane 4 neighborhood function results in the creation of two distinct image objects. The diagonal 8 neighborhood function results in a larger single-image object.

The user specifies the spectral (color) and spatial shape parameters (compactness and smoothness) criteria and the neighborhood function logic (Definiens, 2007). A specially designed heuristic algorithm then applies these criteria to individual pixels in the scene and, in effect, grows homogeneous regions (or, if you like, regions with specified amounts of heterogeneity). Once a segment patch exceeds the user-specified parameters, it stops growing. The final result is a new segmented image consisting of image objects (patches) that contain relatively homogeneous spectral and spatial characteristics (Jensen et al., 2006).

Image Segmentation Classification Examples To appreciate the creation of such image objects (i.e., segments or patches) and how they might be used to classify a remotely sensed image, consider the high spatial resolution ADAR 5000 image of a yacht harbor on Pritchard's Island, SC, in the Ace Basin (Figure 9-36). The data were obtained on September 23, 1999, at a spatial resolution of 0.7×0.7 m in four bands (blue, green, red, and near-infrared). Note the significant amount of spatial information in the dataset including

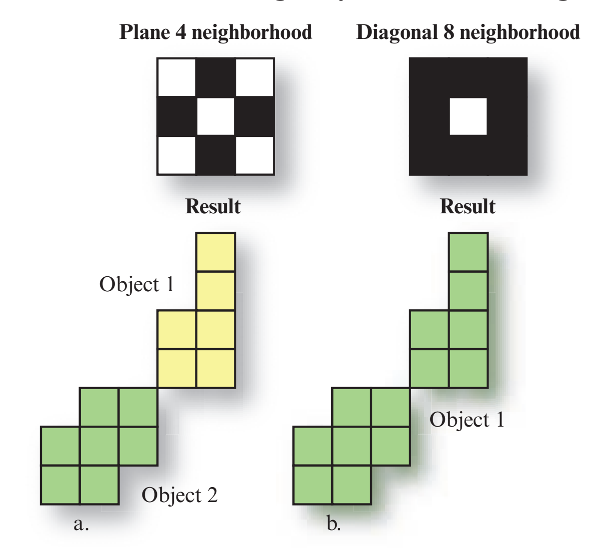


FIGURE 9–35 One of the criteria used to segment a remotely sensed image into image objects is a pixel neighborhood function, which compares an image object being grown with adjacent pixels. The information is used to determine if the adjacent pixel should be merged with the existing image object or be part of a new image object. a) In this example, if a plane 4 neighborhood function is selected, then two image objects would be created because the pixels under investigation are not connected at their plane borders. b) Pixels and objects are defined as neighbors in a diagonal 8 neighborhood if they are connected at a plane border *or* a corner point (Definiens, 2003). In this example, image object 1 can be expanded because it connects at a diagonal corner point. This resulted is a larger image object 1. Other types of neighborhood functions could be used.

building rooftops oriented toward and away from the Sun, automobiles on the concrete or asphalt pavement, upland forest texture, shadows, and Smooth Cordgrass (*Spartina alterniflora*) wetland along the major tidal tributaries.

The *object-based image segmentation process* was weighted so that spectral information (color) was more important than spatial (shape) information (i.e., w_{color} = 0.8). Three spectral bands of ADAR 5000 data (green, red, and near-infrared) were used in the segmentation process. The spatial (shape) parameter was more heavily weighted to smoothness (0.9) than compactness (0.1). The user-specified inputs were then used to identify various "scales" of image segmentation wherein each successive scale file contained larger segmented image objects. This process is commonly referred to as *multiresolution image segmentation* (Baatz and Schape, 2000; Definiens, 2007). Figure 9-36 shows multiresolution image objects associated with four levels of aggregation or scale: 10, 20, 30, and 40. Each image object (i.e., patch, segment, or polygon, if you

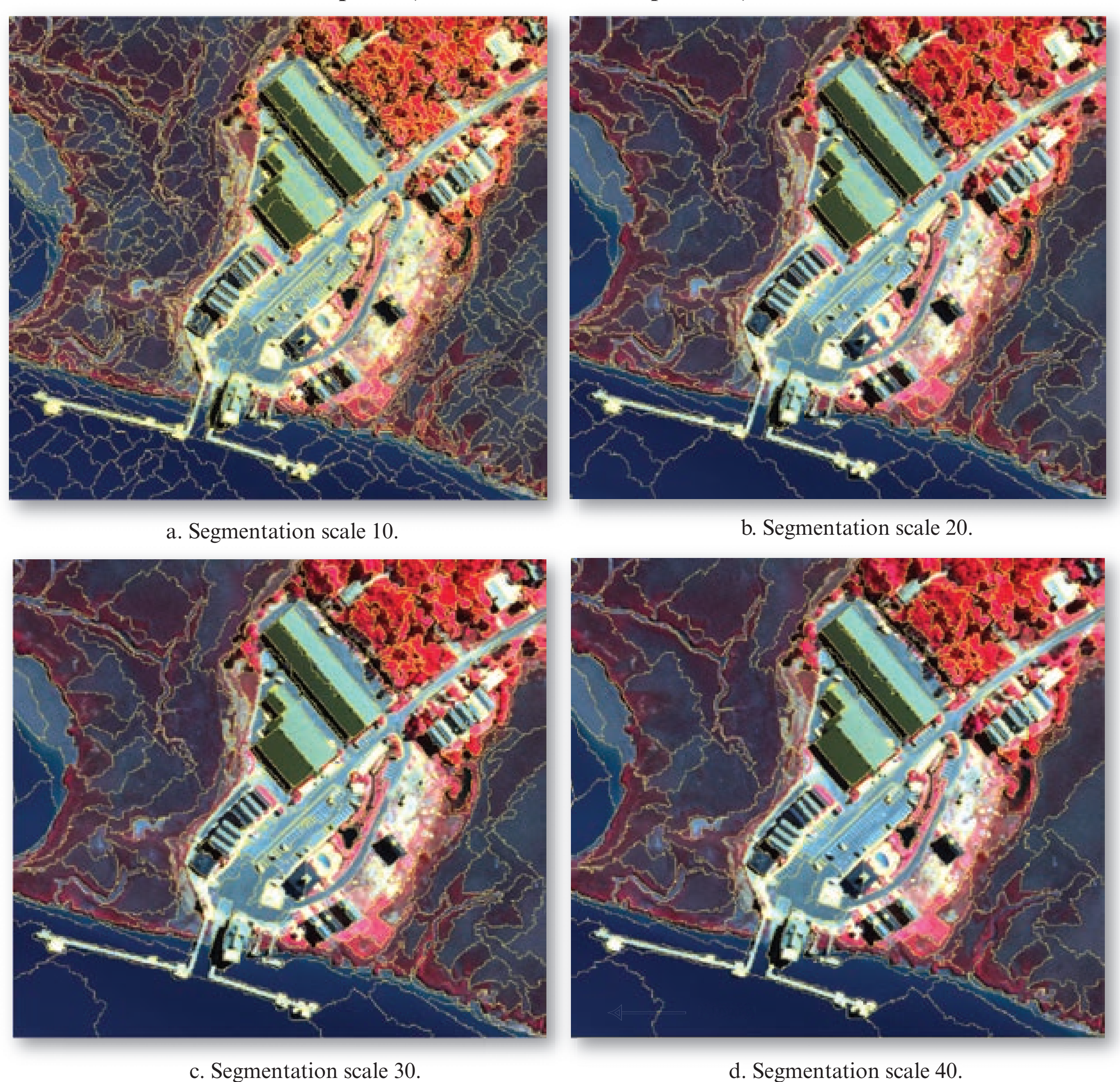


FIGURE 9–36 Color composite of ADAR 5000 imagery of a yacht harbor on Pritchard's Island, SC, in the Ace Basin obtained on September 23, 1999, at a spatial resolution of 0.7 × 0.7 m. Multiresolution image segmentation was performed using three bands (green, red, and near-infrared) at four segmentation scales: 10, 20, 30, and 40. The segmentation process was weighted so that spectral information (color) was more important than spatial (shape) information (weighted 0.8 to 0.2, respectively). The spatial parameter was more heavily weighted to smoothness (0.9) than compactness (0.1).

prefer) in each segmentation scale is hierarchically related to all other files. In other words, the boundaries of one large polygon at segmentation scale 40 may be topologically linked to perhaps four smaller polygons at segmentation scale 10. The four smaller polygons at scale 10 will have exactly the same outer-perimeter coordinates as the larger polygon at scale 40.

One of the most important aspects of performing this type of image segmentation using remote sensor data is that each image object in the dataset (i.e., each polygon) contains not only the mean spectral values of all pixels found within each band but also various spatial measurements that characterize the shape of the polygon. Table 9-12 summarizes several image object metrics that may be computed for each polygon.

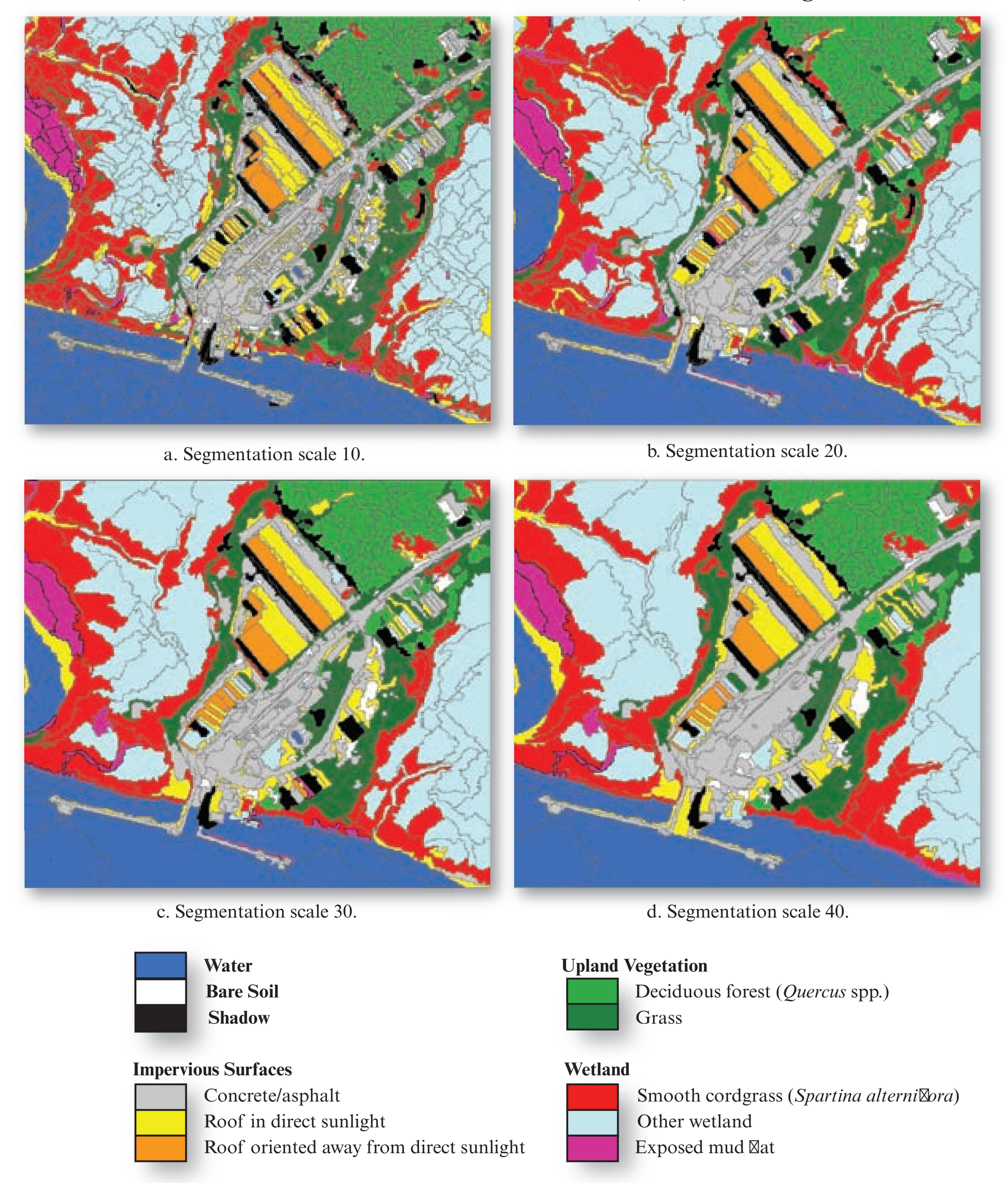


FIGURE 9–37 Four OBIA classification maps derived from four different sets of image segmentation scale data of a yacht harbor on Pritchard's Island, SC, in the Ace Basin.

The *OBIA classification* of a segmented image is substantially different from performing a per-pixel classification (Lu and Weng, 2007; Duro et al., 2012). First, the analyst is not constrained to using just spectral information. He or she may choose to use a) the mean spectral information in conjunction with b) various shape measures associated with each image object (polygon) in the dataset (Definiens, 2007). This introduces flexibility and robustness. Once selected, the spectral and spatial attributes of each polygon may be

TABLE 9–12 A selection of various landscape-ecology metrics that can be computed from individual image objects (patches) that have been segmented within an image (based on O'Neill et al., 1997; Frohn and Hao, 2006; Jensen, 2007).

Image Object Metrics	Algorithm	Description
Mean	$\mu_L = \dfrac{\sum_{i=1}^{n} v_i}{n}$	Image object (i.e., polygon, patch) mean value (μ_L) in a single layer is calculated by summing all the pixel values within the image object (v_i) and dividing by the total number of pixels in the object, *n*. If the image object is composed of spectral brightness values, then $v_i = BV_i$. Otherwise, v_i may be any type of data (e.g., radar backscatter, LIDAR elevation value, DEM elevation value).
Spectral Mean	$b = \dfrac{\sum_{i=1}^{n_L} \mu_i}{N_L}$	For a single image object (i.e., polygon, patch), the sum of all mean layer values (μ_i) divided by the total number of layers (n_L) (i.e., bands).
Spectral Ratio of an Image Object	$\text{ratio}_L = \dfrac{\mu_L}{\sum_{i=1}^{n_L} \mu_i}$	The mean of a single image object (μ_L) divided by the sum of all spectral layers (bands) associated with this image object (μ_i).
Standard Deviation of an Image Object	$\sigma_L = \dfrac{\sum_{i=1}^{n_L} (v_i - \mu_L)^2}{n-1}$	For a single image object (i.e., polygon, patch) in a single layer, the sum of all pixel layer values (v_i) subtracted from the mean of the image object (μ_L) squared, divided by the total number of pixels in the polygon (*n*) minus one.
Mean Difference to Neighboring Image Object	$\Delta c_L = \dfrac{1}{l} \sum_{i=1}^{nn} l_{si}(\mu_i - \mu_{Li})$	For a single image object (i.e., polygon, patch), the mean difference to a direct neighbor is calculated using l = border length in pixels of the image object, l_{si} = border length shared with a direct (attached neighbor), (μ_i) = image object mean value of layer *i*, (μ_{Li}) = neighboring image object mean value of layer *i*, and *nn* = number of neighbors.
Length-to-Width Ratio	$\gamma = \dfrac{l}{w}$	The length-to-width ratio (γ) is computed by dividing the length (l) by the width (w) of an image object. This is approximated by determining the dimensions of the smallest box that bounds the polygon.
Area	$A = \sum_{i=1}^{n} a_i$	For georeferenced data, the area of an image object (*A*) equals the summation of the true area (a_i) of each of the *n* pixels in the image object. An image object consisting of six Landsat Thematic Mapper 30 × 30 m pixels would have an area of 6 pixels × 90m^2 = 540 m^2.
Length	$l = \sqrt{A \cdot \gamma}$	Length of an image object is approximated using the length-to-width ratio. If the image object has a curved shape, it is better to break it into sub-subjects.
Width	$w = \sqrt{\dfrac{A}{\gamma}}$	Width of an image object is approximated using the length-to-width ratio. If the image object has a curved shape, it is better to break it into sub-objects.
Border Length	$bl = \sum_{i=1}^{n} e_i$	The border length of an image object is the sum of the number of edges (e_i) that touch *all* neighboring image objects (or that touch the outside edge of a scene).

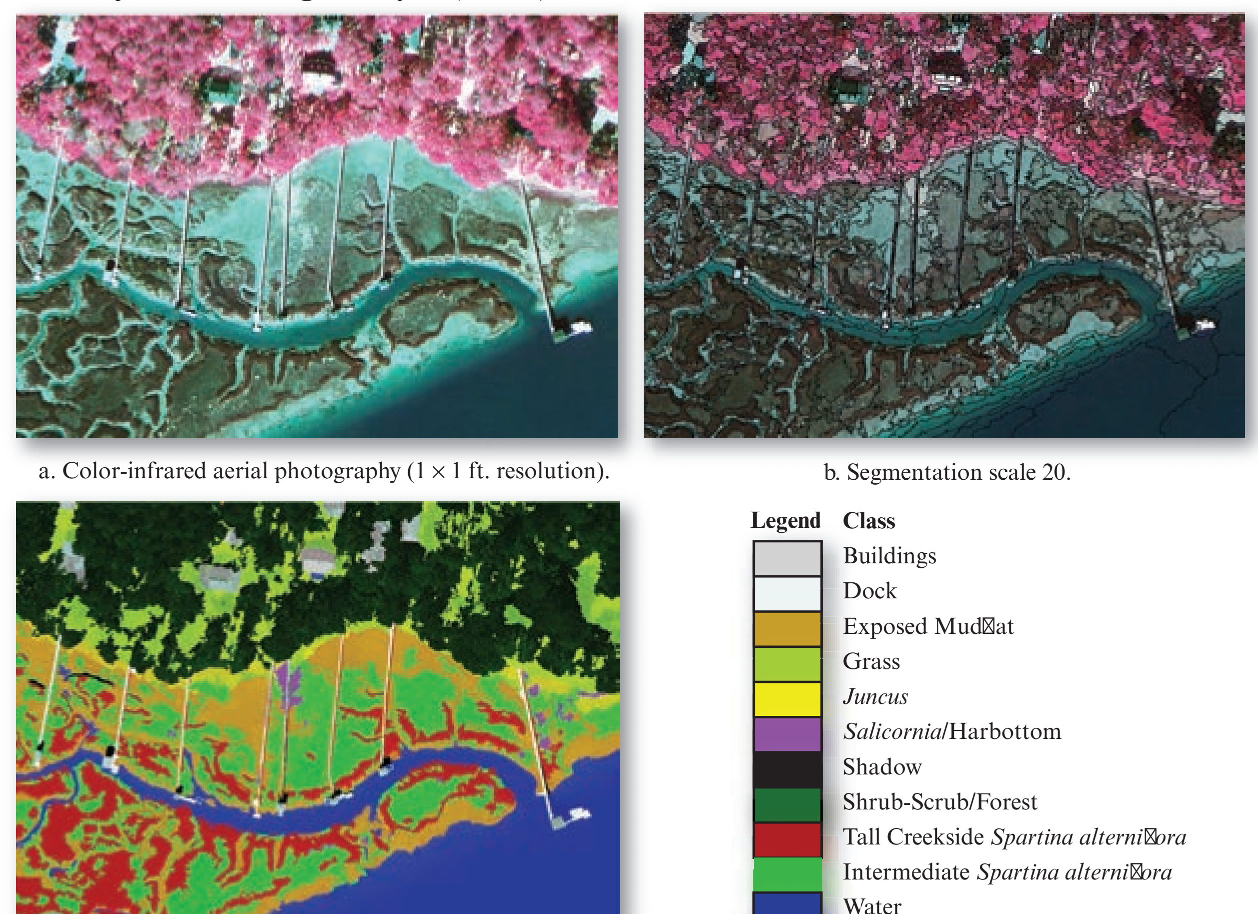
TABLE 9–12 A selection of various landscape-ecology metrics that can be computed from individual image objects (patches) that have been segmented within an image (O'Neill et al., 1997; Frohn and Hao, 2006; Jensen, 2007).

Image Object Metrics	Algorithm	Description
Shape Index	$si = \frac{bl}{4 \times \sqrt{A}}$	The shape index of an image object is the border length (*bl*) divided by four times the square root of the area (*A*). The smoother the shape, the lower the value. The greater the value, the more fractal the shape.
Density	$d = \frac{\sqrt{n}}{1 + \sqrt{Var(X) + Var(Y)}}$	The density of an image object is the area of the object divided by its radius. It is approximated by computing the variance of all *x*- and *y*-coordinates of all *n* pixels forming the image object. Density is a surrogate for compactness. The more compact the object, the higher its density and the more the shape is like a square.
Asymmetry	$k = 1 - \frac{n}{m}$	Image object asymmetry is one minus the ratio of the length of the minor (*n*) and major (*m*) axes of an ellipse enclosing the image object. The greater the asymmetry, the greater the value.
Classified Map Summary Metrics		
Dominance	$D = 1 - \left[\sum_{k=1}^{n} \frac{(-P_k \cdot \ln P_k)}{\ln(n)} \right]$	Dominance ($0 < D < 1$) identifies the extent to which the landscape is dominated by a single land-cover type where $0 < P_k < 1$ is the proportion of land-cover type *k* and *n* is the total number of land-cover types in the land-cover map.
Contagion	$C = 1 - \left[\sum_{i} \sum_{j} \frac{(-P_{ij} \cdot \ln P_{ij})}{2 \ln(n)} \right]$	Contagion ($0 < C < 1$) expresses the probability that land cover is more clumped than the random expectation where P_{ij} is the probability that a pixel of cover type *i* is adjacent to type *j*.
Fractal Dimension	F	The fractal dimension, *F*, of patches (image objects) indicates the extent of human reshaping of the landscape. It is calculated by regressing the log of the patch perimeter against the log of the patch area for each patch on the landscape. The index equals twice the slope of the regression line. Patches < 4 pixels are excluded.

input to a variety of classification algorithms for analysis (e.g., parallelepiped, nearest-neighbor, minimum distance, maximum likelihood).

The classification process is usually very fast because individual image objects (i.e., polygons) rather than individual pixels are assigned to specific classes. Consider the Pritchard's Island example where the dataset consists of 543 rows and 460 columns. A per-pixel classification would process 249,780 pixels. Figure 9-37a depicts the object-oriented classification of just 2,391 polygons into nine classes using segmentation scale 10 data. Only 753 image objects (polygons) were classified in Figure 9-37b (segmentation scale 20 data); 414 in Figure 9-37c (segmentation scale 30 data), and 275 in Figure 9-37d (segmentation scale 40 data).

Examination of Figure 9-37 reveals some interesting aspects of OBIA classification. First, the user must decide what level of polygon aggregation is acceptable for the task at hand. In this example, a classification based on the use of segmentation scale 20 data (Figure 9-37b) is probably the best representation. It provides the greatest amount of accurate information for both very small objects and larger, geographically-extensive surface materials such as forest and wetland. Conversely, if only segmentation scale 40 data (Figure 9-37d) were used, then the classification of the wetland would contain serious error. Perhaps the information that the user wants is best extracted using a combination of information present in image segmentation scale 10 for relatively small rooftop materials and segmentation scale 40 for identifying homogeneous forest areas. The user must decide which of the various segmented files



c. Classification of individual objects.

FIGURE 9–38 a) Digital frame camera color-infrared aerial photography obtained in 2006 at 1 × 1 ft. spatial resolution. b) OBIA image segmentation at a scale of 20. c) Classification of the individual objects according to the U.S. Fish & Wildlife Service (USFS) Cowardin et al. (1979) estuarine intertidal habitat classification scheme summarized in Figures 9-5 and 9-6.

are of most value. He or she can then enter very specific rules into the classification process if desired

A second example provides a more detailed examination of estuarine marine habitat using 2006 high spatial resolution digital camera aerial photography (1 × 1 ft.) for an area near Bluffton, SC (Figure 9-38). After experimentation, a segmentation scale of 20 was judged to be the best level to create useful polygons associated with the various wetland land-cover types. Color was once again more important than spatial (shape) information (i.e., w_{color} = 0.8) and the spatial (shape) parameter was more heavily weighted to smoothness (0.9) than compactness (0.1). The supervised OBIA classification provides detailed wetland information with classes derived from the U.S. Fish & Wildlife Service Cowardin et al. (1979) wetland classification scheme summarized in Figures 9-5 and 9-6.

OBIA Classification Considerations

Once the object-based image segmentation and classification is complete, it may be desirable to evaluate the characteristics of all the patches (image objects) in the map to address important applications. O'Neill (1997) suggested that the functional health of an ecosystem can be monitored if the following three landscape pattern and structure metrics (indices) are monitored through time: dominance, contagion, and fractal dimension (Table 9-12). Other studies use different landscape pattern and structure metrics (Batistella et al., 2003; Frohn and Hao, 2006; Jensen et al., 2006).

Per-pixel classifications often appear pixelated. Conversely, object-oriented classification can appear fractal. The analyst conducting an object-oriented classification should decide what level (scale) of segmentation is most acceptable for the task at hand and

TABLE 9–13 Selected image segmentation freeware (accessible in 2013).

Segmentation Program	Developer	Location
Efficient Graph-Based Image Segmentation	Felzenszwalb, P.F. and D. P. Huttenlocher, 2004; (University of Chicago)	http://www.cs.brown.edu/~pff/segment/.
JSEG - Segmentation of Color-Texture Regions	Deng, Y. and Manjunath, 2001; (University of California at Santa Barbara)	http://vision.ece.ucsb.edu/segmentation/jseg/.
Multiscale Normalized Cuts Segmentation Toolbox for MATLAB	Cour, T., Benezit, F. and J. Shi, 2005; (University of Pennsylvania)	http://www.timotheecour.com/software/ncut_multiscale/ncut_multiscale.html.
SPRING - Georeferenced Information Processing System	Brazilian National Institute for Space Research, 2014	http://www.dpi.inpe.br/spring/english/index. html.

most easily understood by lay viewers when the final thematic map(s) are produced.

Segmentation into image objects allows the use of geographical and landscape ecology concepts involving neighborhood, distance, and location for analyzing remotely sensed data. It also facilitates the merging or fusion of multiple types of remote sensor data for a particular application (Kuehn et al., 2002). Object-based image segmentation and classification is a major paradigm shift when compared with per-pixel classification (Lu and Weng, 2007; Blaschke et al., 2014). It will become increasingly important not only for single-date image classification but also for change detection (e.g., Im, 2006; Im et al., 2008; Campbell, 2010).

For example, Jensen et al. (2006) extracted very accurate agricultural land-cover information using OBIA applied to SPOT multispectral data. The land-cover information was used to model agricultural water demand in South Africa. Hamilton et al. (2007) developed an OBIA-based forest stand delineation system for the U.S. Dept. of Agriculture Forest Service. Hofmann et al. (2008) used OBIA applied to Quickbird data to inventory informal settlements in Rio de Janeiro. Tsai et al. (2011) used OBIA to extract building footprints from high spatial resolution imagery of Accra, Ghana. Myint et al. (2011) found OBIA yielded more accurate results for classifying urban land cover from high spatial resolution imagery than traditional per-pixel classifiers such as maximum likelihood. Duro et al. (2012) compared pixel-based and object-based image analysis approaches for classifying broad land-cover classes over agricultural landscapes. Interestingly, they found no significant difference in the results obtained using OBIA when compared with three pixel-based supervised machine learning algorithms discussed in the next chapter (decision tree, random forest, and the support vector machine).

The following selected commercial digital image processing software has OBIA capability: Definiens *e*Cognition; Intergraph ERDAS Imagine; Trimble ENVI; and Clark Lab IDRISI. Selected sources of image segmentation freeware are summarized in Table 9-13 (updated from Hamilton et al., 2007).



Incorporating Ancillary Data in the Classification Process

An analyst photo-interpreting a color aerial photograph of the terrain often has at his or her disposal 1) systematic knowledge about the soils, geology, vegetation, hydrology, and geography of the area, 2) the ability to visualize and comprehend the landscape's color, texture, height, and shadows, 3) the ability to place much of this diverse information in context to understand site conditions and associations among phenomena, and 4) historical knowledge about the area. Conversely, 95% of all remote sensing digital image classifications attempt to accomplish the same task using a single variable, i.e., an object's spectral reflectance characteristics (color) or black-and-white tone. Therefore, it is not surprising that there is error in remote sensing–derived classification maps. Why should we expect the maps to be extremely accurate when the information provided to the classification algorithm is so rudimentary?

Numerous scientists recognize this condition and have attempted to improve the accuracy and quality of remote sensing–derived land-cover classification by incorporating ancillary data in the classification process (e.g., Hutchinson, 1982; McIver and Friedl, 2002; Lu

and Weng, 2007). *Ancillary data* are any type of spatial or nonspatial information that may be of value in the image classification process, including elevation, slope, aspect, geology, soils, hydrology, transportation networks, political boundaries, and vegetation maps. Ancillary data are not without error. Analysts who want to incorporate ancillary data into the classification process should be aware of several considerations.

Problems Associated with Ancillary Data

First, ancillary data were usually produced for a specific purpose and it was not to improve remote sensing classification accuracy. Second, the nominal, ordinal, or interval thematic attributes on the collateral maps may be inaccurate or incomplete. Great care must be exercised when generalizing the classes found on the ancillary map source materials as we try to make them compatible with the remote sensing investigation classes of interest.

Third, considerable ancillary information is stored in analog map format. The maps must be digitized, translated, rotated, rescaled, and often resampled to bring the dataset into congruence with the remote sensing map projection. During this process the locational attributes of the phenomena may be moved from their true planimetric positions. This assumes that the ancillary data were planimetrically accurate to begin with. Unfortunately, considerable ancillary data were never recorded in their proper planimetric positions. For example, old soil surveys published by the U.S. Soil Conservation Service were compiled onto uncontrolled photomosaics. Analysts trying to use such data must be careful that they do not introduce more error into the classification process than they are attempting to remove.

Approaches to Incorporating Ancillary Data to Improve Remote Sensing Classification Maps

Several approaches may be used to incorporate ancillary data in the image classification process that should improve results. These include incorporating the data before, during, or after classification through geographical stratification, classifier operations, and/or post-classification sorting. Ancillary data may also be incorporated using object-oriented image segmentation, neural networks, expert systems, and decision-tree classifiers.

Geographical Stratification

Ancillary data may be used *prior* to classification to subdivide the regional image into strata, which may then be processed independently. The goal is to increase the homogeneity of the individual stratified image datasets to be classified. For example, what if we wanted to locate spruce fir in the Colorado Rockies but often encountered misclassification up and down the mountainside? One approach would be to stratify the scene into just two files: one with elevations from 0 to 2,600 ft above sea level (dataset 1) and another with elevation > 2,600 ft ASL (dataset 2). We would then classify the two datasets independently. Spruce fir do not grow below 2,600 ft ASL; therefore, during the classification process we would not label *any* of the pixels in dataset 1 spruce fir. This would keep spruce fir pixels from being assigned to forested areas that cannot ecologically support them. Errors of commission for spruce fir should be reduced when datasets 1 and 2 are put back together to compile the final map and are compared to a traditional classification. If specific ecological principles are known, the analyst could stratify the area further using slope and aspect criteria to refine the classification.

Stratification is a conceptually simple tool and, carefully used, can be effective in improving classification accuracy. Illogical stratification can have severe implications. For example, differences in training set selection for individual strata and/or the vagaries of clustering algorithms, if used, may produce different spectral classes on either side of strata boundaries. Edge-matching problems become apparent when the final classification map is put together from the maps derived from the individual strata.

Classifier Operations

Several methods may be used to incorporate ancillary data during the image classification process. One of the most useful is the logical channel method. A *per-pixel logical channel* classification includes ancillary data as one of the channels (features) used by the classification algorithm. For example, a dataset might consist of three IKONOS bands of spectral data plus two additional bands (percent slope and aspect) derived from a digital elevation model. The entire five-band dataset is acted on by the classification algorithm. When ancillary data are incorporated into traditional classification algorithms as logical channels, the full range of information available in the ancillary data is used (e.g., Ricchetti, 2000). With logical channel addition, ancillary data are given equal weight to single spectral bands unless weights are assigned in a maximum likelihood classifier (Lawrence and Wright, 2001). Chen and Stow (2003) used the logical channel approach to classify land cover using multiple types of imagery obtained at different spatial resolutions.

The context of a pixel refers to its spatial relationship with any other pixel or group of pixels throughout the

scene. *Contextual logical channel* classification occurs when information about the neighboring (surrounding) pixels is used as one of the features in the classification. *Texture* is one simple contextual measure that may be extracted from an $n \times n$ window (see Chapter 8) and then added to the original image dataset prior to classification (Stow et al., 2003). It is important to remember that contextual information may also be derived from *non*image ancillary sources, such as maps showing proximity to roads, streams, and so on.

A second approach involves the use of *a priori* probabilities in the classification algorithm. The analyst gets the *a priori* probabilities by evaluating historical summaries of the region (e.g., last year cotton accounted for 80% of the acreage, hay 15%, and barley 5%). These statistics can be incorporated directly into a maximum likelihood classification algorithm as weights $p(w_i)$ to the classes, as previously discussed. Prior probabilities can improve classification results by helping to resolve confusion among classes that are difficult to separate and by reducing bias when the training sample is not representative of the population being classified. McIver and Friedl (2002) point out that the use of *a priori* probabilities in maximum likelihood classification is often problematic in practice. They developed a useful method for incorporating prior probabilities into a nonparametric decision-tree classifier.

Another approach involves image segmentation, as discussed in the previous section. This method can incorporate both spectral and nonspectral ancillary data, which are subjected to multiresolution segmentation to produce polygons (patches) that contain relatively uniform spectral and spatial characteristics. This is a straightforward way of merging both spectral and nonspectral information.

Ancillary data have been incorporated into modern classification methods such as expert systems and neural networks (Stow et al., 2003; Qiu and Jensen, 2004). These approaches incorporate the ancillary data directly into the classification algorithms and are usually not dependent on *a priori* weights. Chapter 10 describes how such systems work and the ease with which ancillary data may be introduced, including the fact that they are not confined to normally distributed data.

Machine-learning approaches have been used to establish rule-based classification systems where expert knowledge was inadequate (Huang and Jensen, 1997; Myint et al., 2011). Lawrence and Wright (2001) used rule-based classification systems based on classification and regression tree (CART) analysis to incorporate ancillary data into the classification process.

Post-Classification Sorting

This method involves the application of very specific rules to (1) initial remote sensing classification results and (2) spatially distributed ancillary information. For example, Hutchinson (1982) classified Landsat MSS data of a desert area in California into nine initial classes. He then registered slope and aspect maps derived from a digital elevation model with the classification map and applied 20 if–then rules to the datasets (e.g., if the pixel was initially classified as an active sand dune and if the slope <1%, then the pixel is a dry lake bed). This eliminated confusion between several of the more prominent classes in this region (e.g., between the bright surfaces of a dry lake bed [playa] and the steep sunny slopes of large sand dunes). Similarly, Cibula and Nyquist (1987) used post-classification sorting to improve the classification of Landsat MSS data for Olympic National Park. Topographic (elevation, slope, and aspect) and watershed boundary data (precipitation and temperature) were analyzed in conjunction with the initial land-cover classification using Boolean logic. The result was a 21-class forest map that was just as accurate as the initial map but contained much more information.

The incorporation of ancillary data in the remote sensing classification process is important. However, the choice of variables to include is critical. Common sense suggests that the analyst thoughtfully select only variables with conceptual and practical significance to the classification problem at hand. Incorporating illogical or suspect ancillary information can rapidly consume limited data analysis resources and lead to inaccurate results.



References

American Planning Association, 2014a, *Land-Based Classification Standard*, Washington: American Planning Association, http://www.planning.org/lbcs/.

American Planning Association, 2014b, *Land-Based Classification Standards: LBCS Tables*, Washington: American Planning Association, https://www.planning.org/lbcs/standards/pdf/InOneFile.pdf, 168 p.

Anderson, J. R., Hardy, E., Roach, J., and R. Witmer, 1976, *A Land-Use and Land-Cover Classification System for Use with Remote Sensor Data*, Washington: U.S. Geological Survey, Professional Paper #964, 28 p.

Baatz, M., and A. Schape, 2000, "Multiresolution Segmentation: An Optimization Approach for High Quality Multiscale Image Segmentation," in Strobl, J., Blaschke, T., and G. Griesebner (Eds.), *Angewandte Geographische Informationsverarbeitung XII*, Heidelberg: Wichmann, 12–23.

Baatz, M., Benz, U., Dehghani, S., Heymen, M., Holtje, A., Hofmann, P., Ligenfelder, I., Mimler, M., Sohlbach, M., Weber, M., and G. Willhauck, 2001, *eCognition User Guide*, Munich: Definiens Imaging GmbH, 310 p.

Batistella, M., Robeson, S., and E. F. Moran, 2003, "Settlement Design, Forest Fragmentation, and Landscape Change in Rondonia, Amazonia," *Photogrammetric Engineering & Remote Sensing*, 69(7):805–812.

Beauchemin, M., and K. B. Fung, 2001, "On Statistical Band Selection for Image Visualization," *Photogrammetric Engineering & Remote Sensing*, 67(5):571–574.

Belward, A. S., and J. Skoien, 2014, "Who Launched What, When and Why: Trends in Global Land-Cover Observation Capacity from Civilian Earth Observation Satellites," *ISPRS Journal of Photogrammetry & Remote Sensing*, in press.

Benz, U., 2001, "Definiens Imaging GmbH: Object-Oriented Classification and Feature Detection," *IEEE Geoscience and Remote Sensing Society Newsletter*, (Sept.), 16–20.

Blaschke, T., 2010, "Object Based Image Analysis for Remote Sensing," *ISPRS Journal of Photogrammetry & Remote Sensing*, 65:2–16.

Blaschke, T., and J. Strobl, 2001, "What's Wrong with Pixels? Some Recent Developments Interfacing Remote Sensing and GIS," *GIS*, Heidelberg: Huthig GmbH & Co., 6:12–17.

Blaschke, T., and 10 co-authors, 2014, "Geographic Object-based Image Analysis – Towards a New Paradigm," *ISPRS Journal of Photogrammetry & Remote Sensing*, 87:180–191.

Botkin, D. B., Estes, J. E., MacDonald, R. B., and M. V. Wilson, 1984, "Studying the Earth's Vegetation from Space," *Bioscience*, 34(8):508–514.

Brazilian National Institute for Space Research, 2014, *SPRING Georeferenced Information Processing System*, http://www.dpi.inpe.br/spring/english/index.html.

Campbell, J. B., 2010, "Chapter 19: Information Extraction from Remotely Sensed Data," in Bossler, J. D., Campbell, J. B., McMaster, R. B. and C. Rizos (Eds.), *Manual of Geospatial Science and Technology*, 2nd Ed., New York: Taylor & Francis, 363–390.

Campbell, J. B., and R. H. Wynne, 2011, *Introduction to Remote Sensing*, 5th Ed., New York: Guilford, 684 p.

Chen, D., and D. Stow, 2003, "Strategies for Integrating Information from Multiple Spatial Resolutions into Land-use/Land-cover Classification Routines," *Photogrammetric Engineering & Remote Sensing*, 69(11):1279–1287.

Cibula, W. G., and M. O. Nyquist, 1987, "Use of Topographic and Climatological Models in a Geographical Data Base to Improve Landsat MSS Classification for Olympic National Park," *Photogrammetric Engineering & Remote Sensing*, 53:67–75.

Colditz, R. R., Schmidt, M., Conrad, C., Hansen, M. C., and S. Dech, 2011, "Land Cover Classification with Coarse Spatial Resolution Data to Derive Continuous and Discrete Maps for Complex Regions," *Remote Sensing of Environment*, 115:3264–3275.

Colomina, I., and P. Molina, 2014, "Unmanned Aerial Systems for Photogrammetry & Remote Sensing: A Review," *ISPRS Journal of Photogrammetry & Remote Sensing*, 92:79–97.

Congalton, R. G., and K. Green, 2009, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd E., Boca Raton, FL: Lewis Publishers, 183 p.

Cour, T., Benezit, F., and J. Shi, 2005, "Spectral Segmentation with Multiscale Graph Decomposition," *IEEE International Conference on Computer Vision and Pattern Recognition 2005 (CVPR)*, 2: 1124–1131.

Cowardin, L. M., Carter, V., Golet, F. C., and E. T. LaRoe, 1979, *Classification of Wetlands and Deepwater Habitats of the United States*, Washington: U.S. Fish & Wildlife Service, FWS/ OBS-79/31, 103 p.

Dalponte, M., Bruzzone, L., Vescovo, L., and D. Gianelle, 2009, "Role of Spectral Resolution and Classifier Complexity in the Analysis of Hyperspectral Images of Forest Areas," *Remote Sensing of Environment*, 113:2345–2355.

Definiens, 2003, *eCognition Professional*, Munich: Definiens-imaging.com.

Definiens, 2007, *eCognition Developer 7 Reference Book*, Munich: Definiens AG, 195 p.

Detenbeck, N. E., 2002, *Methods for Evaluating Wetland Condition: #7 Wetlands Classification*, Washington: Environmental Protection Agency, Report #EPA-822-R-02-017, 43 p.

Dronova, I., Gong, P., and L. Wang, 2011, "Object-based Analysis and Change Detection of Major Wetland Cover Types and their Classification Uncertainty during the Low Water Period at Poyang Lake, China," *Remote Sensing of Environment*, 115:3220–3236.

Duro, D. C., Franklin, S. E., and M. G. Dube, 2012, "A Comparison of Pixel-based and Object-based Image Analysis with Selected Machine Learning Algorithms for the Classification of Agricultural Landscapes using SPOT-5 HRG Imagery," *Remote Sensing of Environment*, 118:259–272.

Duda, R. O., Hart, P. E., and D. G. Stork, 2001, *Pattern Classification*, New York: John Wiley & Sons, 654 p.

EPA, 2008, *Nutrient Criteria Technical Guidance Manual - Wetlands*, EPA Report #EPA-822-B-08-001, 25 p.

ERDAS, 2013, *ERDAS Field Guide*, Atlanta: Intergraph, Inc., 772 p.

Exelis ENVI, 2013, *ENVI* Classic Tutorial: *Classification Methods*, Boulder: Exelis Visual Information Solutions, 25 p., http://www.exelisvis.com/.

Felzenszwalb, P. F., and D. P. Huttenlocher, 2004, "Efficient Graph-Based Image Segmentation" *International Journal of Computer Vision*, 59(2):1–26.

Ferro, C. J., and T. A. Warner, 2002, "Scale and Texture in Digital Image Classification," *Photogrammetric Engineering & Remote Sensing*, 68(1):51–63.

FGDC, 2014, *Natural Vegetation Classification Standard*, Washington: Federal Geographic Data Committee, http://www.fgdc.gov/standards/projects/FGDC-standards-projects/vegetation/index_html/?searchterm=vegetation.

FGDC Wetland Mapping Standard, 2014, Wetland Mapping Standard, http://www.fgdc.gov/standards/projects/FGDC-standards-projects/wetlands-mapping/index.html.

Foody, G. M., 2002, "Status of Land Cover Classification Accuracy Assessment," *Remote Sensing of Environment*, 80:185–201.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramakkutty, N., Sibley, A., and X. Huang, 2010, "MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of New Datasets," *Remote Sensing of Environment*, 114:168–182.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and C. Schaaf, 2002, "Global Land Cover Mapping from MODIS: Algorithms and Early Results," *Remote Sensing of Environment*, 83:287–302.

Frohn, R. C., and Y. Hao, 2006, "Landscape Metric Performance in Analyzing Two Decades of Deforestation in the Amazon Basin of Rondonia, Brazil," *Remote Sensing of Environment*, 100:237–251.

Fry, J. A., et al., 2011, "National Land Cover Database for the Conterminous United States," *Photogrammetric Engineering & Remote Sensing*, 2011(9):859–864.

Fuchs, H., Magdon, P., Kleinn, C., and H. Flessa, 2009, "Estimating Aboveground Carbon in a Catchment of the Siberian Forest Tundra: Combining Satellite Imagery and Field Inventory," *Remote Sensing of Environment*, 113:518–531.

Giri, C., Zhu, Z., and B. Reed, 2005, "A Comparative Analysis of the Global Land Cove 2000 and MODIS Land Cover Data Sets," *Remote Sensing of Environment,* 94:123–132.

Hamilton, R., Megown, K., Mellin, T., and I. Fox, 2007. *Guide to Automated Stand Delineation using Image Segmentation,* RSAC-0094-RPT1. Salt Lake City: U.S. Department of Agriculture, Forest Service, Remote Sensing Applications Center. 16 p.

Hardin, P. J., 1994, "Parametric and Nearest-neighbor Methods for Hybrid Classification: A Comparison of Pixel Assignment Accuracy," *Photogrammetric Engineering & Remote Sensing*, 60(12):1439–1448.

Herold, M., Guenther, S., and K. C. Clarke, 2003, "Mapping Urban Areas in the Santa Barbara South Coast using IKONOS and *e*Cognition," *eCognition Application Note*, Munich: Definiens ImgbH, 4(1):2 p.

Herold, M., Scepan, J., and K. C. Clarke, 2002, "The Use of Remote Sensing and Landscape Metrics to Describe Structures and Changes in Urban Land Uses," *Environment and Planning A*, 34:1443–1458.

Hodgson, M. E., 1988, "Reducing the Computational Requirements of the Minimum-distance Classifier," *Remote Sensing of Environment*, 25:117–128.

Hodgson, M. E., and R. W. Plews, 1989, "*N*-dimensional Display of Cluster Means in Feature Space," *Photogrammetric Engineering & Remote Sensing*, 55(5):613–619.

Hofmann, P, Strobl, J., Blaschke, T., and H. J. Kux, 2008, "Detecting Informal Settlements from QuickBird Data in Rio de Janeiro Using an Object-based Approach," in *Object-based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications,* Blaschke, T., Lang, S. and G. J. Hay, Eds., Springer: Berlin, 531–554.

Homer, C. H., Fry, J. A., and C. A. Barnes, 2012, *The National Land Cover Database*, U.S. Geological Survey Fact Sheet 2012–3020, 4 p.

Huang, K., 2002, "A Synergistic Automatic Clustering Technique for Multispectral Image Analysis," *Photogrammetric Engineering & Remote Sensing*, 68(1):33–40.

Huang, X., and J. R. Jensen, 1997, "A Machine Learning Approach to Automated Construction of Knowledge Bases for Image Analysis Expert Systems That Incorporate GIS Data," *Photogrammetric Engineering & Remote Sensing*, 63(10):1185–1194.

Hutchinson, C. F., 1982, "Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement," *Photogrammetric Engineering & Remote Sensing*, 48(1):123–130.

Ifarraguerri, A., and M. W. Prairie, 2004, "Visual Method for Spectral Band Selection," *IEEE Geoscience and Remote Sensing Letters*, 1(2):101–106.

IGBP-IHDP, 2002, *Land Use and Land Cover Change (LUCC): A Joint IGBP- IHDP Research Project*, http://www.igbp.net/.

Im, J., 2006, *A Remote Sensing Change Detection System Based on Neighborhood/Object Correlation Image Analysis, Expert Systems, and an Automated Calibration Model*, Ph.D. dissertation, Dept. of Geography, University of South Carolina, Columbia, SC.

Im, J., Jensen, J. R., and J.A. Tullis, 2008, "Object-based Change Detection Using Correlation Image Analysis and Image Segmentation Techniques," *International Journal of Remote Sensing*, 29(2): 399–423.

Im, J., Jensen, J. R., Jensen, R. R., Gladden, J., Waugh, J., and M. Serrato, 2012, "Vegetation Cover Analysis of Hazardous Waste Sites in Utah and Arizona Using Hyperspectral Remote Sensing," *Remote Sensing*, 2012(4):327–353.

Jensen J. R., 2007, *Remote Sensing of The Environment: An Earth Resource Perspective*, 2nd Ed., Upper Saddle River, NJ: Prentice-Hall, 592 p.

Jensen, J. R., and D. C. Cowen, 1999, "Remote Sensing of Urban/Suburban Infrastructure and Socioeconomic Attri-

butes," *Photogrammetric Engineering & Remote Sensing*, 65:611–622.

Jensen, J. R., and D. L. Toll, 1982, "Detecting Residential Land Use Development at the Urban Fringe," *Photogrammetric Engineering & Remote Sensing*, 48:629–643.

Jensen, J. R., and M. E. Hodgson, 2011, *Predicting the Impact of Sea Level Rise in the Upper May River near Bluffton, SC, using an Improved LiDAR-derived Digital Elevation Model, Land Cover Extracted from High Resolution Digital Aerial Imagery, and Sea Level Rise Scenarios*, Charleston: The Nature Conservancy, 78 p.

Jensen, J. R., and R. R. Jensen, 2013, *Introductory Geographic Information Systems*, Upper Saddle River: Pearson, Inc., 400 p.

Jensen, J. R., Im, J., Hardin, P., and R. R. Jensen, 2009, "Chapter 19: Image Classification," in *The Sage Handbook of Remote Sensing*, Warner, T. A., Nellis, M. D. and G. M. Foody, (Eds.), 269–296.

Jensen, J. R., Botchway, K., Brennan-Galvin, E., Johannsen, C., Juma, C., Mabogunje, A., Miller, R., Price, K., Reining, P., Skole, D., Stancioff, A., and D. R. F. Taylor, 2002, *Down to Earth: Geographic Information for Sustainable Development in Africa*, Washington: National Research Council, 155 p.

Jensen, J. R., Garcia-Quijano, M., Hadley, B., Im, J., Wang, Z., Nel, A. L., Teixeira, E., and B. A. Davis, 2006, "Remote Sensing Agricultural Crop Type for Sustainable Development in South Africa," *Geocarto International*, 21(2):5–18.

Jia, K., and 7 co-authors, 2014, "Land Cover Classification of Finer Resolution Remote Sensing Data Integrating Temporal Features from Time Series Coarser Resolution Data," *ISPRS Journal of Photogrammetry & Remote Sensing*, 93:49–55.

Kerr, J., and M. Ostrovsky, 2003, "From Space to Species: Ecological Applications for Remote Sensing," *Trends in Ecology and Evolution*, 18(6):299–305.

Konecny, G., 2014, *Geoinformation: Remote Sensing, Photogrammetry and Geographic Information Systems*, Boca Raton: CRC Press, 416 p.

Kuehn, S., Benz, U., and J. Hurley, 2002, "Efficient Flood Monitoring Based on RADARSAT-1 Images Data and Information Fusion with Object-Oriented Technology," *Proceedings, IGARSS,* 3 p.

Lawrence, R. L., and A. Wright, 2001, "Rule-Based Classification Systems Using Classification and Regression Tree (CART) Analysis," *Photogrammetric Engineering & Remote Sensing,* 67(10):1137–1142.

Lillesand, T., Kiefer, R., and J. Chipman, 2008, *Remote Sensing and Image Interpretation*, 6th Ed., New York: John Wiley, 756 p.

Liu, K., and F. Xia, 2010, "Assessing Object-based Classification,: Advantages and Limitations," *Remote Sensing Letters*, 1(4):187–194.

Lo, C. P., and A. K. Yeung, 2007, *Concepts and Techniques of Geographic Information Systems*, 2nd Ed., Upper Saddle River, NJ: Prentice-Hall, 492 p.

Lu, D., and Q. Weng, 2007, "A Survey of Image Classification Methods and Techniques for Improving Classification Performance," *International Journal of Remote Sensing*, 28, 823–870.

Mahlein, A., Rumpf, T., Welke, P., Dehne, H., Plumer, L., Steiner, U., and E. Oerke, 2013, "Development of Spectral Indices for Detecting and Identifying Plant Diseases," *Remote Sensing of Environment*, 1128:21–30.

Mausel, P. W., Kamber, W. J., and J. K. Lee, 1990, "Optimum Band Selection for Supervised Classification of Multispectral Data," *Photogrammetric Engineering & Remote Sensing*, 56(1):55–60.

McIver D. K., and M. A. Friedl, 2002, "Using Prior Probabilities in Decision-tree Classification of Remotely Sensed Data," *Remote Sensing of Environment*, 81:253–261.

Memarsadeghi, N., Mount, D. M., Netanyau, N., and J. Le Mogne, 2007, "A Fast Implementation of the ISODATA Clustering Algorithm," *International Journal of Computational Geometry & Applications*, 17(1):71–103.

MRLC, 2014 *National Land Cover Database 2006 Product Legend*, Washington: MRLC, http://www.mrlc.gov/nlcd06_leg.php.

Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., and Q. Weng, 2011, "Per-pixel vs. Object-based Classification of Urban Land Cover Extraction using High Spatial Resolution Imagery," *Remote Sensing of Environment*, 115:1145–1161.

NASA Earth Observatory, 2002, "NASA's *Terra* Satellite Refines Map of Global Land Cover," *Earth Observatory News*, 02-126, August 13, 2002.

NASA GSFC, 2014, *MODIS Land Cover and Land Cover Change Data Product #12*, Greenbelt, MD: NASA Goddard Space Flight Center, http://modis.gsfc.nasa.gov/data/dataprod/dataproducts.php?MOD_NUMBER=12.

NOAA C-CAP, 2014, *Coastal Change Analysis Program (C-CAP) Regional Land Cover*, Charleston: NOAA Coastal Services Center, http://www.csc.noaa.gov/digitalcoast/data/ccapregional.

O'Neill, R. V., Hunsaker, C. T., Jones, K. B., Ritters, K. H., Wickham, J. D., Schwarz, P., Goodman, I. A., Jackson, B., and W. S. Bailargeon, 1997, "Monitoring Environmental Quality at the Landscape Scale," *BioScience*, 47(8):513–519.

Ozkan, C., and F. S. Erbek, 2003, "The Comparison of Activation Functions for Multispectral Landsat TM Image Classification," *Photogrammetric Engineering & Remote Sensing*, 69(11):1225–1234.

Pasher, J., and D. J. King, 2010, "Multivariate Forest Structure Modelling and Mapping using High Resolution Air-

borne Imagery and Topographic Information," *Remote Sensing of Environment,* 114:1718–1732.

Pena-Barragan, J. M., Ngugi, M. K., Plant, R. E., and J. Six, 2011, "Object-based Crop Classification using Multiple Vegetation Indices, Textural features and Crop Phenology," *Remote Sensing of Environment*, 115:1301–1316.

Phillips, T., et al., 2011, "Modeling Moulin Distribution on Sermeq Avannarleq Glacier using ASTER and WorldView Imagery and Fuzzy Set Theory," *Remote Sensing of Environment*, 115:2292–2301.

Pu, R., Gong, P., Michishita, R., and T. Sasagawa, 2008, "Spectral Mixture Analysis for Mapping Abundance of Urban Surface Components from the Terra/ASTER Data," *Remote Sensing of Environment*, 112:949–954.

Qiu, F., and J. R. Jensen, 2004, "Opening the Neural Network Black Box and Breaking the Knowledge Acquisition Bottleneck of Fuzzy Systems for Remote Sensing Image Classification," *International Journal of Remote Sensing*, 25(9):1749–1768.

Ricchetti, E., 2000, "Multispectral Satellite Image and Ancillary Data Integration for Geological Classification," *Photogrammetric Engineering & Remote Sensing,* 66(4):429–435.

Rich, R. L., Frelich, L., Reich, P. B., and M. E. Bauer, 2010, "Detecting Wind Disturbance Severity and Canopy Heterogeneity in Boreal Forest by Coupling High-Spatial Resolution Satellite Imagery and Field Data," *Remote Sensing of Environment*, 114:299–308.

Richards, J. A., 2013, *Remote Sensing Digital Image Analysis*, 5th Ed., NY: Springer, 494 p.

Rocchini, D., Ricotta, C., and A. Chiarucci, 2007, "Using Satellite Imagery to Assess Plant Species Richness: The Role of Multispectral Systems," *Applied Vegetation Science*, 10:325–331.

Schowengerdt, R. A., 2007, *Remote Sensing: Models and Methods for Image Processing*, 3rd Ed., San Diego, CA: Academic Press, 515 p.

Song, C., 2005, "Spectral Mixture Analysis for Subpixel Vegetation Fractions in the Urban Environment: How to Incorporate Endmember Variability?" *Remote Sensing of Environment*, 95:248–263.

Stow, D., Coulter, L., Kaiser, J., Hope, A., Service, D., Schutte, K., and A. Walters, 2003, "Irrigated Vegetating Assessments for Urban Environments," *Photogrammetric Engineering & Remote Sensing*, 69(4):381–390.

Swain, P. H. and S. M. Davis, 1978, *Remote Sensing: The Quantitative Approach*, New York: McGraw-Hill, 166–174.

Textron Systems, 2014, *Feature Analyst,* Missoula, MT: Visual Learning Systems, http://www.textronsystems.com/capabilities/geospatial.

Townshend, J. R. G., Huang, C., Kalluri, S., DeFries, R., Liang, S., and K. Yang, 2000, "Beware of Per-pixel Characterization of Land Cover," *International Journal of Remote Sensing*, 21(4):839–843.

Trimble, Inc., 2014, *eCognition*, Version 3.0, Parsippany: Definiens, Inc., http://www.ecognition.com/.

Tsai, Y. H., Stow, D., and J. Weeks, 2011, "Comparison of Object-Based Image Analysis Approaches to Mapping New Buildings in Accra, Ghana Using Multi-Temporal QuickBird Satellite Imagery," *Remote Sensing*, 2011(3):2707–2726.

Wang, F., 1990a, "Improving Remote Sensing Image Analysis through Fuzzy Information Representation," *Photogrammetric Engineering & Remote Sensing*, 56(8):1163–1169.

Wang, F., 1990b, "Fuzzy Supervised Classification of Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, 28(2):194–201.

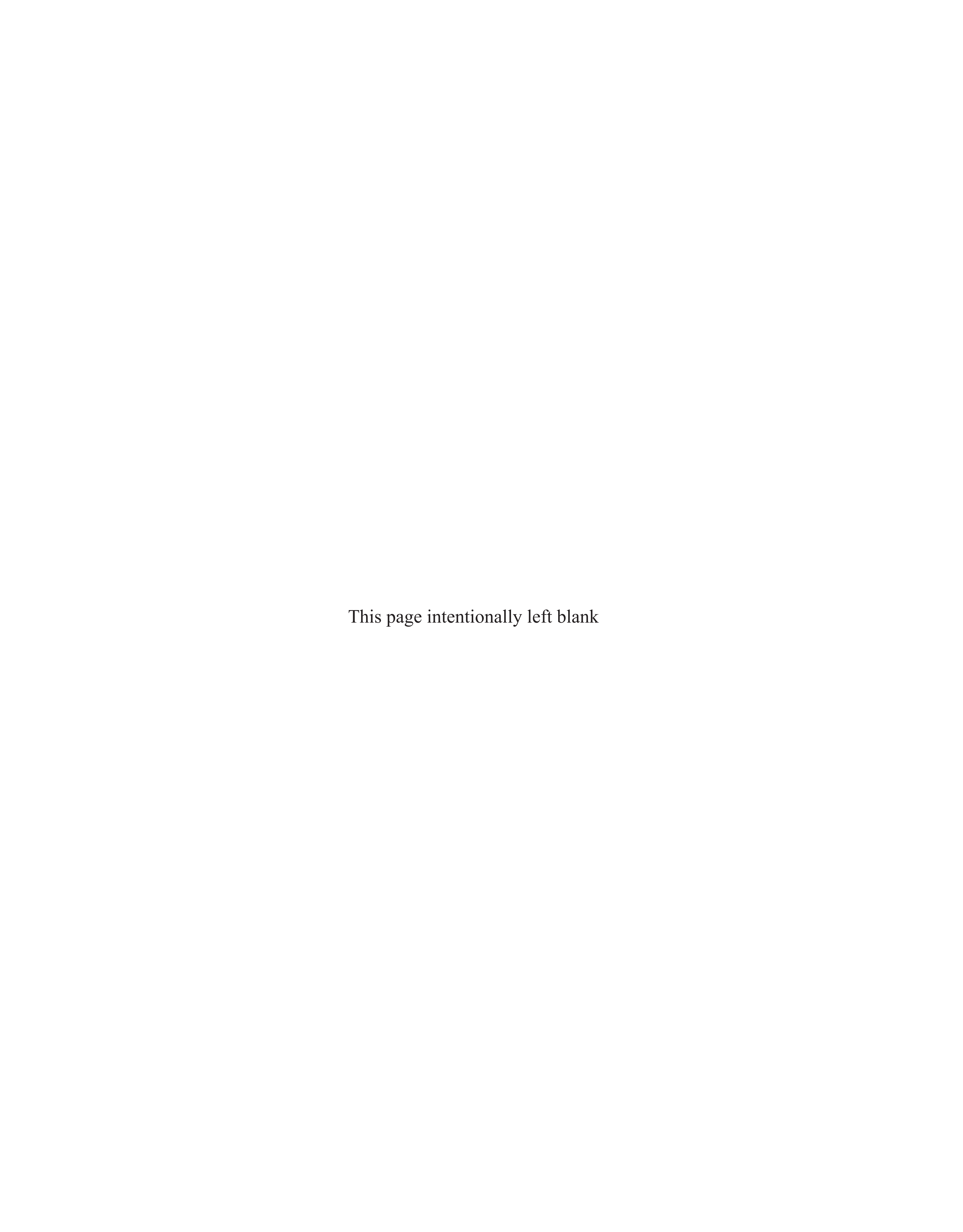
Wang, Z., Jensen, J. R., and J. Im, 2010, "An Automatic Region-based Image Segmentation Algorithm for Remote Sensing Applications," *Environmental Modelling & Software*, 25(10):1149–1165.

Warner, T. A., Almutairi, A., and J. Y. Lee, 2009, "Chapter 33: Remote Sensing of Land Cover Change," in *The Sage Handbook of Remote Sensing*, Warner, T. A., Nellis, M. D., and G. M. Foody, (Eds.), Los Angeles: Sage, 459–472.

Welch, R. A., 1982, "Spatial Resolution Requirements for Urban Studies," *International Journal of Remote Sensing*, 3:139–146.

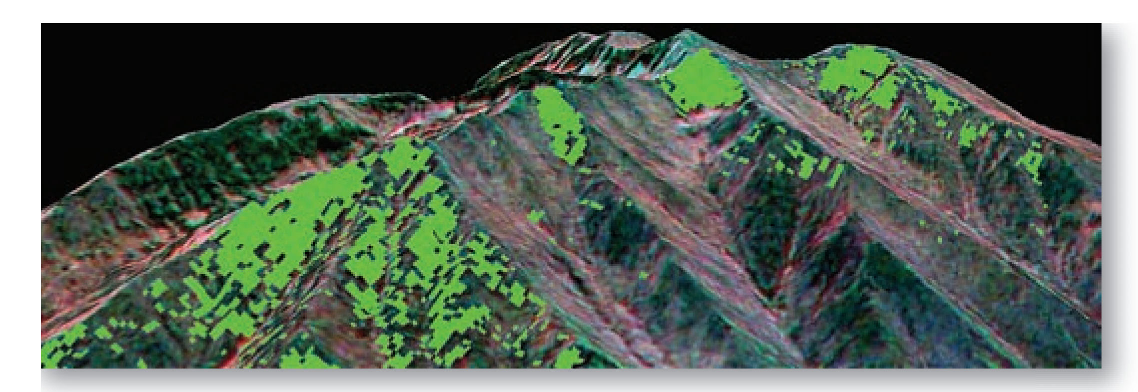
Weng, Q., 2014, *Global Urban Monitoring and Assessment through Earth Observation*, New York: Routledge, 440 p.

Xian, G., Homer, C., and J. Fry, 2009, "Updating the 2001 National Land Cover Database Land Cover Classification using Landsat Imagery Change Detection Methods," *Remote Sensing of Environment*, 113:1133–1147.



This page intentionally left blank

10 INFORMATION EXTRACTION USING ARTIFICIAL INTELLIGENCE



One definition of *artificial intelligence* (AI) is:

> the theory and development of computer systems able to perform tasks normally requiring human intelligence (Oxford, 2014).

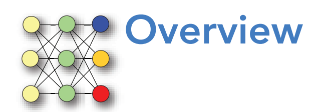
But how do we know when an artificially intelligent system has been created? Ideally we could use the *Turing test,* which suggests that if we are unable to distinguish a computer's response to a problem of interest from a human's response to the same problem, then the computer system is said to have *intelligence* (Turing, 1950). The test or task is for an artificial intelligence program to have a blind conversation with an interrogator for 5 minutes. The interrogator has to guess if the conversation is with an artificial intelligence program or with a real person. The AI program passes the test if it fools the interrogator 30% of the time. Unfortunately, it is very difficult for most artificial intelligence systems to pass the Turing test. For this reason, "the field of AI as a whole has paid little attention to Turing tests," preferring instead to forge ahead developing artificial intelligence applications that simply work (Russell and Norvig, 2010).

Artificial intelligence research was initiated in 1955 when Allen Newell and Herbert Simon at the RAND Corporation proved that computers could do more than calculate.

> They demonstrated that computers were physical symbol systems whose symbols could be made to stand for anything, including features of the real world, and whose programs could be used as rules for relating these features. In this way computers could be used to simulate certain important aspects of intelligence. Thus, the information-processing model of the mind was born (Dreyfus and Dreyfus, 2001).

Unfortunately, artificial intelligence was oversold in the 1960s much like remote sensing was oversold in the 1970s. General artificial intelligence problem solving was found to be much more difficult than originally suggested. Scientists had difficulty getting computers to solve problems that were routinely solved by human experts. Therefore, scientists instead started to investigate the development of artificial intelligence applications in "micro-worlds," or very narrow topical areas. This led to the creation of the first useful artificial intelligence systems for select applications, e.g., games, disease diagnosis (MYCIN), spectrograph analysis (DENDRAL). MYCIN was developed at Stanford University in 1976 to aid physicians in diagnosing and treating patients with infectious blood diseases caused by bacteria in the blood and meningitis. These diseases can be fatal if not recognized and treated quickly. The DENDRAL program solved the problem of inferring molecular structure from the information provided by a mass spectrometer (Buchanan and Lederberg, 1971). DENDRAL was the first truly successful knowledge-intensive system deriving its expertise from large numbers of special-purpose rules. The field advanced at a rapid rate. For example, NASA's REMOTE AGENT program became the first on-board autonomous planning program to control the scheduling of operations for a spacecraft travelling a hundred million miles from Earth (Jonsson et al., 2000). Such expert systems are based on the use of knowledge or rules derived from human experts. The knowledge is extracted from the expert by a knowledge engineer and turned into useful

rule-based reasoning that can, hopefully, be performed by a computer.



Overview

This chapter introduces the fundamental characteristics of several knowledge-based expert systems including machine-learning derived decision trees and regression trees, random forest decision-tree analysis and support vector machines. It then describes the characteristics of neural networks and their advantages and limitations when used to extract information from remotely sense images.



Expert Systems

A knowledge-based *expert system* may be defined as "a system that uses human knowledge to solve problems that normally would require human intelligence" (PC AI, 2002). It uses artificial intelligence techniques in problem-solving processes to support human decision-making, learning, and action (Akerkar and Sajja, 2009). It is the ability to solve problems efficiently and effectively in a relatively narrow problem area and to perform at the level of an expert. Expert systems represent the expert's domain (i.e., subject matter) knowledge base as data and rules within the computer. The rules and data can be called upon when needed to solve problems. A completely different problem within the domain of the knowledge base can be solved without having to reprogram.

Knowledge-based expert systems are used extensively in remote sensing research and production. Machine-learning *decision-tree classifiers* may be used to extract categorical information such as land use/land cover. For example, decision-tree classification is used to create the NASA global MODIS Collection 5 Land Cover type product (Friedl et al., 2010). Both Pena-Barragan et al. (2011) and Duro et al. (2012) used various decision trees to classify agriculture land cover. Machine-learning *regression trees* can also be used to predict continuous quantitative information. For example, Coops et al. (2006) used classification and regression tree (CART) analysis to map the probability of Mountain Pine Beetle infestation in Canada. Im et al. (2012ab) used regression trees to predict LAI and percent impervious surface.

Expert systems may also be used to detect change in complex heterogeneous environments (Yang and Chung, 2002). For example, Im and Jensen (2005) introduced local correlation image analysis for change detection and created neighborhood correlation images (NCIs) associated with bi-temporal high spatial resolution datasets. They used a machine learning decision-tree classifier (C5.0) to extract change information from the bi-temporal data. The USGS uses decision-tree and regression-tree classifiers to detect change associated with its National Land Cover Database (NLCD) products (Xian et al., 2009, 2011).

A knowledge-based expert system consists of the components shown in Figure 10-1, including:

- human expert,
- user interface,
- knowledge base (rule-based domain),
- inference engine,
- on-line databases, and
- user.

It is useful to review the characteristics of each of these components.

Expert System User Interface

The expert system user interface should be easy to use, interactive, and interesting. It should also be intelligent and accumulate user preferences in an attempt to provide the most pleasing communication environment possible. Figure 10-2 depicts a commercially available Knowledge Engineer interface that can be used to develop remote sensing–assisted expert systems. This expert system *shell* was built using object-oriented programming and is easy to use. All of the hypotheses, rules, and conditions for an entire expert system may be viewed and queried from the single user interface.

Creating the Knowledge Base

Images, books, articles, manuals, and periodicals have a tremendous amount of information in them. Practical experience in the field with vegetation, soils, rocks, water, atmosphere, and urban infrastructure is also tremendously important. However, a human must comprehend the information and experiences and turn it into knowledge for it to be useful. Many human beings have trouble interpreting and understanding the information in images, books, articles, manuals, and periodicals. Similarly, some do not obtain much knowledge from field work. Fortunately, some laypersons and scientists are particularly adept at processing their knowledge using three different problem-solving approaches:

- algorithms using conventional computer programs,
- heuristic knowledge-based expert systems, and
- artificial neural networks.

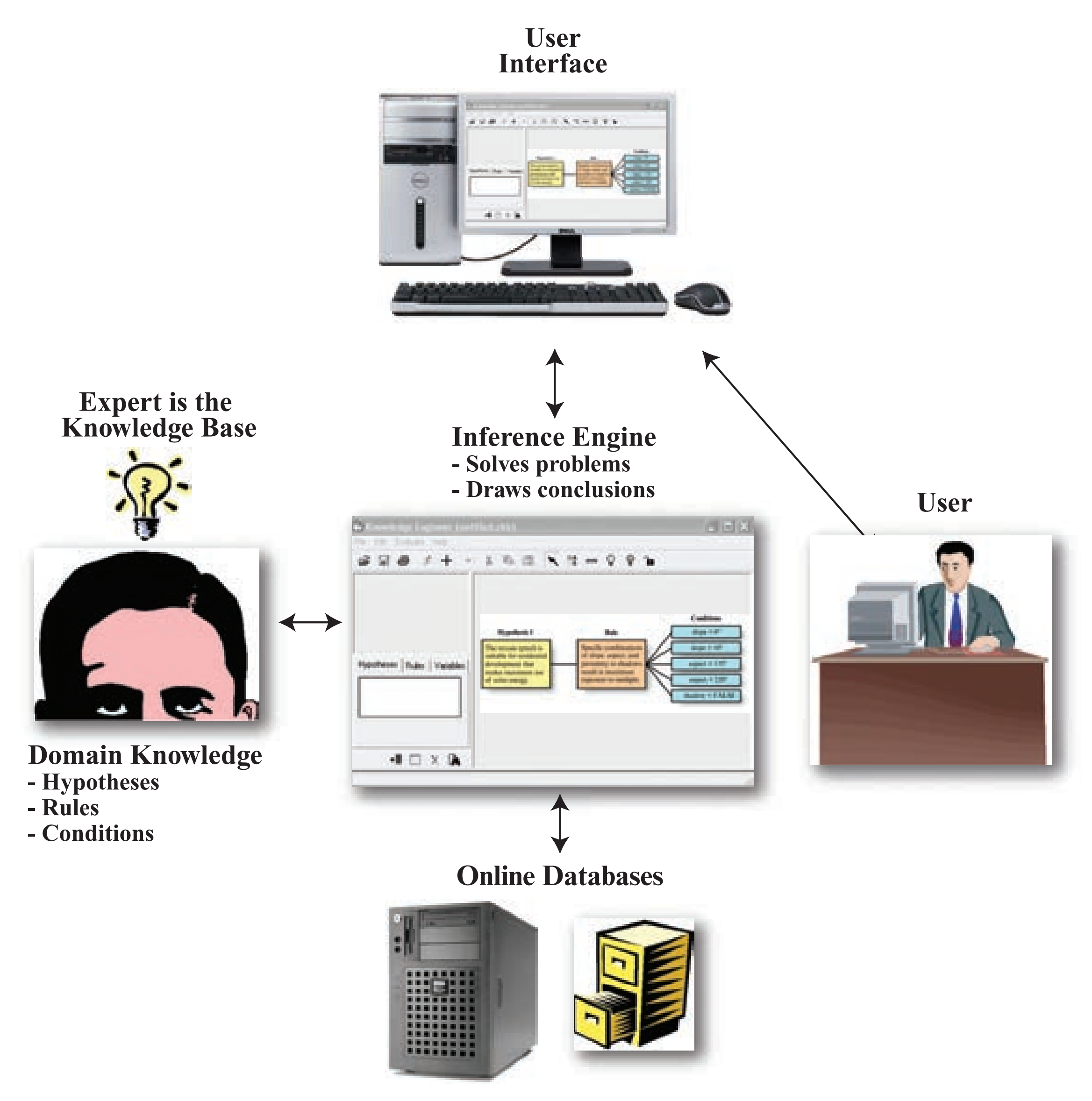


FIGURE 10–1 The components of a typical rule-based expert system. The domain (thematic) knowledge contained in an expert's mind is extracted in the form of a *knowledge base* that consists of hypotheses (problems), rules, and conditions that satisfy the rules. A *user interface* and an *inference engine* are used to encode the knowledge base rules, extract the required information from *online databases*, and solve problems. Hopefully, the information is of value to the *user* who queries the expert system.

Algorithmic Approaches to Problem Solving

Conventional algorithmic computer programs contain little knowledge other than the basic algorithm for solving a specific problem, the necessary boundary conditions, and data. The knowledge is usually embedded in the programming code. As new knowledge becomes available, the program has to be changed and recompiled (Table 10-1).

Heuristic Knowledge-Based Expert System Approaches to Problem Solving

Knowledge-based expert systems, on the other hand, collect many small fragments of human know-how for a specific application area (domain) and place them in a *knowledge base* that is used to reason through a problem, using the knowledge that is most appropriate. Important characteristics that distinguish knowledge-based expert systems from conventional algorithmic systems are summarized in Table 10-1. *Heuristic knowledge* may be defined as "involving or serving as an aid to learning, discovery, or problem solving by experimental and especially by trial-and-error methods. Heuristic computer programs often utilize exploratory problem-solving and self-educating techniques (as the evaluation of feedback) to improve performance" (Merriam-Webster, 2003).

The Problem with Experts Unfortunately, most experts really do not know *exactly* how they perform

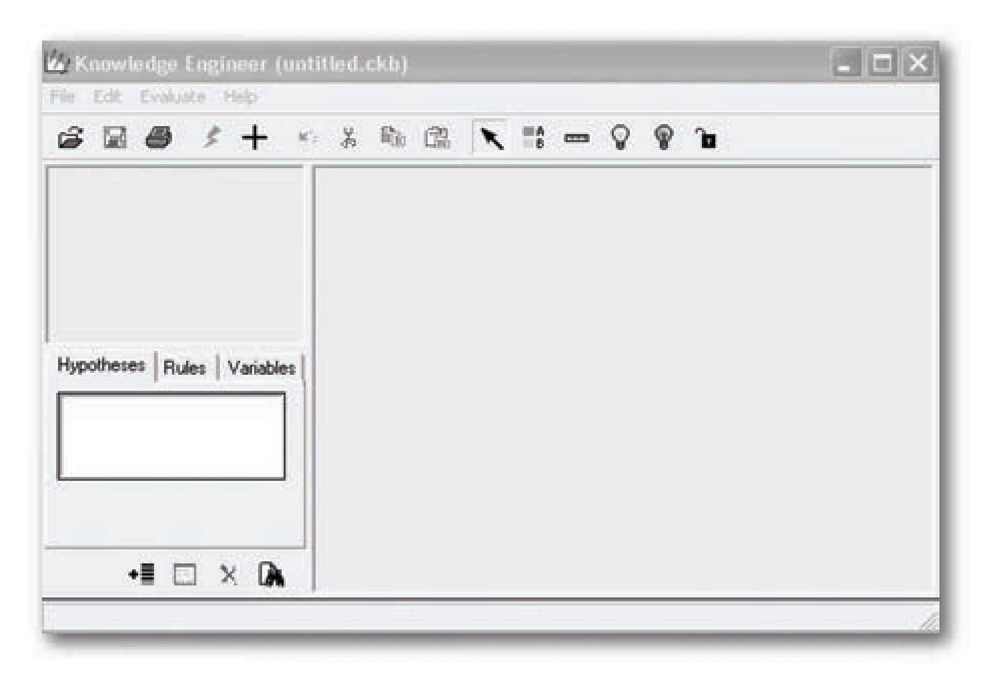


FIGURE 10–2 The Knowledge Engineer interface used in ERDAS Imagine's Expert Classifier (interface courtesy of Hexagon Geospatial).

their expert work (Dreyfus and Dreyfus, 2001). Much of their expertise is derived from experiencing life and observing hundreds or even thousands of case studies. It is often difficult for experts to understand the intricate workings of complex systems much less be able to break them down into their constituent parts and then mimic the decision-making process of the human mind. Therefore, how is it possible to get the knowledge embedded in the mind of an expert into the formal rules and conditions necessary to create an expert system to solve relatively narrowly defined hypotheses (i.e., problems)? This is the primary responsibility of the *knowledge engineer.*

The knowledge engineer interrogates the domain expert and extracts as many rules and conditions as possible that are relevant to the hypotheses (problems) being examined. Ideally, the knowledge engineer has unique capabilities that allow him or her to help build the most appropriate rules. This is not easy. The knowledge engineering process can be costly and time-consuming.

Recently, it has become acceptable for a domain expert (e.g., biologist, geographer, agronomist, forester) to create his or her own knowledge-based expert system by querying oneself and hopefully accurately specifying the rules associated with the problem at hand, for example, using Hexagon's ERDAS Imagine expert system Knowledge Engineer (Hexagon ERDAS Imagine, 2014). When this activity takes place, the expert must have a wealth of knowledge in a certain domain and the ability to formulate a hypothesis and parse the rules and conditions into understandable elements that are very amenable to the "knowledge representation process."

The Knowledge Representation Process

The *knowledge representation process* normally involves encoding information from verbal descriptions, rules of thumb, images, books, maps, charts, tables, graphs, equations, etc. Hopefully, the knowledge base contains sufficient high-quality rules to solve the problem under investigation. *Rules* are normally expressed in the form of one or more "IF condition THEN action" statements. The *condition* portion of a rule statement is usually a fact, e.g., the pixel under investigation must reflect > 45% of the incident near-infrared energy. When certain rules are applied, various operations may take place such as adding a newly derived derivative fact to the database or firing another rule. Rules can be implicit (slope is high) or explicit (e.g., slope > 70%). It is possible to chain together rules, e.g., IF *c* THEN *d*; IF *d* THEN *e*; therefore IF *c* THEN *e*. It is also possible to attach confidences (e.g., 80% confident) to facts and rules. For example, a typical rule used by the MYCIN expert system is (Darlington, 1996):

IF the stain of the organism is gram-negative
  AND the morphology of the organism is rod
    AND the aerobicity of the organism is anaerobic
THEN there is strong suggestive evidence (0.8) that the class of the organism is *Enterobacteriaceae.*

Following the same format, a typical remote sensing rule might be:

TABLE 10–1 Characteristics that distinguish knowledge-based expert systems from conventional algorithmic problem-solving systems (based on Darlington, 1996).

Characteristics	Knowledge-Based Expert System	Conventional Algorithmic System
Paradigm	**Heuristic**. Based on "rules of thumb". The solution steps are implicit (not determined by the programmer). The "solution" is not always correct. It is a declarative problem-solving paradigm.	**Algorithmic**. Solution steps are explicitly written by the programmer. The correct answer is usually obtained. It is a procedural problem-solving paradigm.
Method of Operation	**Reasons with symbols**. Infers the conclusions from known premises. Inference engine decides the order in which the premises are evaluated.	**Computation**. Processes numeric data by sorting, calculating, and computing to obtain an answer.
Processing Unit	**Knowledge**. Usually represented in the form of rules and conditions. Knowledge is active in that an expert system can reason with the knowledge provided to infer new knowledge.	**Data**. Usually represented in arrays or records in languages like C^{++}. Data are typically passive in that they do not give rise to further generations of data.
Control Mechanism	**Inference engine** is usually separate from the domain knowledge.	Data or information and control are usually integrated.
Fundamental Components	**Expert system** = inference + knowledge.	**Conventional algorithmic system** = algorithm(s) + data.
Explanation Capability	**Yes.** An explicit trace of the chain of steps underlying the reasoning processes allows the user to find out how the expert system arrived at its conclusion or why the system is asking for an answer to a particular question.	**No**.

IF blue reflectance is (Condition) <15%
AND green reflectance is (Condition) < 25%
AND red reflectance is (Condition) < 15%
AND near-infrared reflectance is (Condition)> 45%
THEN there is strong suggestive evidence (0.8) that the pixel is vegetated.

Decision Trees The best way to conceptualize an expert system is to use a *decision-tree structure* where rules and conditions are evaluated in order to test hypotheses (Figure 10-3). When decision trees are organized with hypotheses, rules, and conditions, each hypothesis may be thought of as the trunk of a tree, each rule a limb of a tree, and each condition a leaf. This is commonly referred to as a *hierarchical decision-tree classifier* (e.g., Swain and Hauska, 1977; Jensen, 1978; DeFries and Chan, 2000; Stow et al., 2003; Zhang and Wang, 2003; Friedl et al., 2010). The purpose of using a hierarchical structure for labeling objects is to gain a more comprehensive understanding of relationships among objects at different scales of observation or at different levels of detail (Tso and Mather, 2001).

A decision tree takes as input an object or situation described by a set of attributes and returns a decision. The input attributes can be discrete or continuous. The output value can also be discrete or continuous. Learning a discrete-valued function is called *classification learning*. However, many applications require the extraction of biophysical information (e.g., LAI) about a pixel or patch of pixels. Such classification is based on the use of continuous functions and is called *regression learning* (Lawrence and Wright, 2001; Jensen et al., 2009; Im et al., 2012a). We will concentrate on Boolean classification wherein each example is classified as true (positive) or false (negative). A decision tree reaches its decision by performing a sequence of tests (Russell and Norvig, 2010; Im et al., 2012b).

Decision trees predict class membership by recursively partitioning a dataset into more homogeneous subsets (DeFries and Chan, 2000; Im et al., 2012ab). We will first use a simple hypothetical GIS-related example to demonstrate the logic that should be adopted when building the knowledge domain.

Identify the hypothesis: The expert in charge of creating the knowledge domain identifies a hypothesis (problem) to be addressed. This may be a formal hypothesis to be tested using inductive logic and confidence levels or an informal hypothesis that is in search of a logical conclusion:

- ***Hypothesis 1:*** the terrain (pixel) is suitable for residential development that makes maximum use of solar energy (i.e., I will be able to put solar panels on my roof).

Specify the expert system rules: Heuristic rules that the expert has learned over time are the heart and soul of an expert system. If the expert's heuristic rules of thumb are indeed based on correct principles, then the expert system will most likely function properly. If the expert does not understand all the subtle nuances of

Building Blocks of an Expert System

Conditions

Rules

Hypothesis

Hypothesis

FIGURE 10–3 The fundamental building blocks of an expert system include hypotheses (problems), rules, and conditions. The rules and conditions operate on data (information). It is possible to address more than one hypothesis using an expert system.

the problem, has left out important variables or interaction among variables, or applied too much significance (weight) to certain variables, the expert system outcome may not be accurate. Therefore, the creation of accurate, definitive rules is extremely important (Hodgson et al., 2003). Each rule provides the specific conditions to accept the hypothesis to which it belongs. For example, a single rule that might be associated with hypothesis 1 is:

- specific combinations of terrain slope, aspect, and proximity to shadows result in maximum exposure to sunlight.

Specify the rule conditions: The expert would then specify one or more conditions that must be met for each rule. For example, conditions for the rule stated above might include:

- slope > 0º, AND
- slope < 10º (i.e., the terrain should ideally lie on terrain with 1º to 9º slope), AND
- aspect > 135º, AND
- aspect < 220º (i.e., in the Northern Hemisphere the terrain should ideally face south between 136º and 219º to obtain maximum exposure to sunlight), AND
- the terrain is not intersected by shadows cast by neighboring terrain, trees, or other buildings (derived from a viewshed model).

In this case, the hierarchical decision-tree diagram would look like Figure 10-4.

To function properly, the expert system must have access to the data (variables) required by the rules. In this case three types of spatial information would be required for one rule: 1) slope, 2) aspect, and 3) a shadow viewshed file derived from a detailed digital elevation model.

Inference Engine

The terms *reasoning* and *inference* are generally used to describe any process by which conclusions are reached (Russell and Norvig, 2010). Thus, the hypotheses, rules, and conditions are passed to the *inference engine* where the expert system is implemented. One or more conditional statements within each rule are evaluated using the spatial data (e.g., 135º < aspect < 220º). Multiple conditions within a rule are evaluated based on Boolean AND logic. While all of the conditions within a rule must be met to satisfy the rule, any single rule within a hypothesis can cause that hypothesis to be accepted or rejected. In some cases, rules within a hypothesis disagree on the outcome and a decision must be made using rule confidences (e.g., a confidence of 0.8 in a preferred rule and a confidence of 0.7 in another) or the order of the rules (e.g., preference given to the first rule) as the factor. The confidences and order associated with the rules are normally stipulated by the expert.

The inference engine interprets the rules in the knowledge base to draw conclusions. The inference engine may use backward- and/or forward-chaining strategies. Both backward and forward inference processes consist of a chain of steps that can be traced by the expert

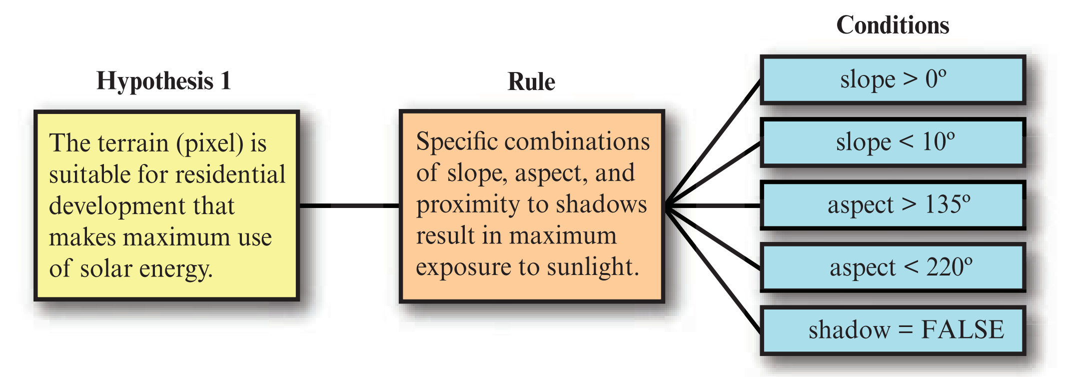


FIGURE 10–4 A human-derived decision-tree expert system with a rule and conditions to be investigated by an inference engine to test hypothesis 1.

system. This enables expert systems to explain their reasoning processes, which is an important and positive characteristic of expert systems. You would expect a doctor to explain how he or she came to a certain diagnosis regarding your health. An expert system can provide explicit information about how a particular conclusion (diagnosis) was reached.

An expert system *shell* provides a customizable inference engine. Expert system shells come equipped with an inference mechanism (backward chaining, forward chaining, or both) and require knowledge to be entered according to a specified format. Expert system shells qualify as languages, although certainly with a narrower range of application than most programming languages (PC AI, 2002). Typical artificial intelligence programming languages include LISP, developed in the 1950s, PROLOG, developed in the 1970s, and now object-oriented languages such as C^{++}.

On-Line Databases

The rules and conditions may be applied and evaluated using data and/or information stored in on-line databases. The databases can take a variety of forms. It can be spatial and consist of remotely sensed images and thematic maps in raster and vector format. However, the database may also consist of charts, graphs, algorithms, pictures, and text that are considered important by the expert. The database should contain detailed, standardized metadata.

Expert Systems Applied to Remote Sensor Data

The use of expert systems in remote sensing research will be demonstrated using two different methodologies used to create the rules and conditions in the knowledge base. The first expert system classification is based on the use of formal rules developed by a human expert. The second example involves expert system rules not generated by humans, but derived automatically by an inductive machine-learning algorithm based on training data that is input by humans into the system. Both methods will be used to identify white fir forest stands on Maple Mountain in Utah County, UT, using Landsat Enhanced Thematic Mapper Plus (ETM^+) imagery and several topographic variables extracted from a digital elevation model of the area. The comparison will demonstrate important characteristics of human expert versus machine-learning rule development strategies.



Decision-Tree Classification Based on Human-Derived Rules

This example is based on rules specified by a human expert to map white fir (*Abies concolor*) forest on Maple Mountain located in Utah County, UT. Maple Mountain rises from the 5,000 ft. valley floor to 10,200 ft. above sea level (ASL). A Landsat ETM^+ panchromatic image of the mountain and valley obtained on August 10, 1999, is shown in Figure 10-5a. A color-composite of bands 4, 3, and 2 (NIR, red, green) is shown in Figure 10-5b. A Normalized Difference Vegetation Index (NDVI) image derived from the Landsat ETM^+ red and near-infrared bands (4 and 2) is displayed in Figure 10-5c.

The goal of this exercise is to accurately extract the spatial distribution of forest land-cover information using the original remotely sensed data, image-derived products (e.g., NDVI), elevation from a digital elevation model (DEM), and DEM-derived products (e.g., slope and aspect). A U.S. Geological Survey 30 × 30 m digital elevation model is shown in Figure 10-5d. Contour, shaded relief, slope, and aspect were extracted from the digital elevation model (Figures 10-5e through i).

TABLE 10–2 A hypothesis (class), variables, and conditions necessary to extract white fir (*Abies concolor*) forest cover information from Maple Mountain, UT, using remote sensing and digital elevation model data. The Boolean logic with which these variables and conditions are organized within a chain of inference may be controlled by the use of rules and sub-hypotheses.

Hypothesis	Variables	Conditions
White fir (*Abies concolor*)	**Aspect** **Elevation** **Slope** **Multispectral**	**Aspect** = 300° to 45° **Elevation** >1,200 m **Slope** = 25° to 50° **Remote sensing reflectance** TM band 1 Blue = 44 to 52 TM band 2 Green = 31 to 40 TM band 3 Red = 22 to 32 TM band 4 Near-infrared = 30 to 86 TM band 5 Mid-infrared = 19 to 47 **NDVI** = 0.2 to 0.7

Hypotheses to Be Tested

The hypothesis (class) to be tested (extracted) from the spatial data was white fir (*Abies concolor*). Many other types of land cover are present in the scene but we will focus our attention on white fir. The structure of the expert system logic is displayed in decision-tree format in the expert system interface shown in Figure 10-6. Note that the hypothesis represents the base of the decision tree (laid on its side).

Rules (Variables)

A human expert who worked for the Uinta National Forest developed the knowledge base (hypotheses, rules, and conditions) to identify white fir habitat from all other land cover in this area. The rules and conditions were based on remote sensing multispectral reflectance characteristics and derivatives (e.g., NDVI), elevation above sea level, and the expert's knowledge of microclimate and soil moisture conditions that are controlled primarily by terrain slope and aspect (Table 10-2).

Conditions

The expert identified very specific conditions that are associated with the remote sensing reflectance data, elevation, slope and aspect. This part of Utah is in a semiarid mountain and range province and receives relatively little moisture in the summer months. Therefore, south-facing slopes usually have reduced soil moisture conditions throughout much of the summer. The expert knows that white fir requires substantial year-round soil moisture. Therefore, it thrives best on north-facing slopes ($300^{o} \leq$ aspect $\leq 45^{o}$). White fir favors elevations > 1,900 m ASL. It out-competes other forest types when the slope is between 25° and 50°.

White fir forest physiology causes it to have relatively distinct red and near-infrared reflectance characteristics when compared with other vegetation types on the mountain. This translates into useful normalized difference vegetation index (NDVI) values,

$$NDVI = \frac{\rho_{TM4} - \rho_{TM3}}{\rho_{TM4} + \rho_{TM3}}, \qquad (10.1)$$

based on Landsat 7 ETM$^+$ band 3 (red) and band 4 (near-infrared) reflectance data. The specific multispectral remote sensing conditions for each band and for the NDVI are summarized in Table 10-2. The hypothesis, rules, and conditions specified in the expert system interface are summarized in Figure 10-6.

Inference Engine

The hypothesis, rules, and conditions created by the expert were passed to the inference engine. The inference engine processed the rules and conditions in conjunction with the required spatial data. This produced the classification map depicting the spatial distribution of white fir displayed in planimetric map format in Figure 10-7e and draped over the Landsat ETM$^+$ color composite and digital elevation model in Figure 10-7f.

Unfortunately, this classification took quite a bit of research on the part of the human expert to come up with the expert system rules and conditions. It would be much better if there were some way for the expert system to generate its own rules based on training data. This leads us to a discussion of machine learning.



Classification Based on Machine Learning Decision Trees and Regression Trees

The heart of an expert system is its knowledge base. The usual method of acquiring knowledge in a computer-usable format to build a knowledge base involves human domain experts and knowledge engineers, as

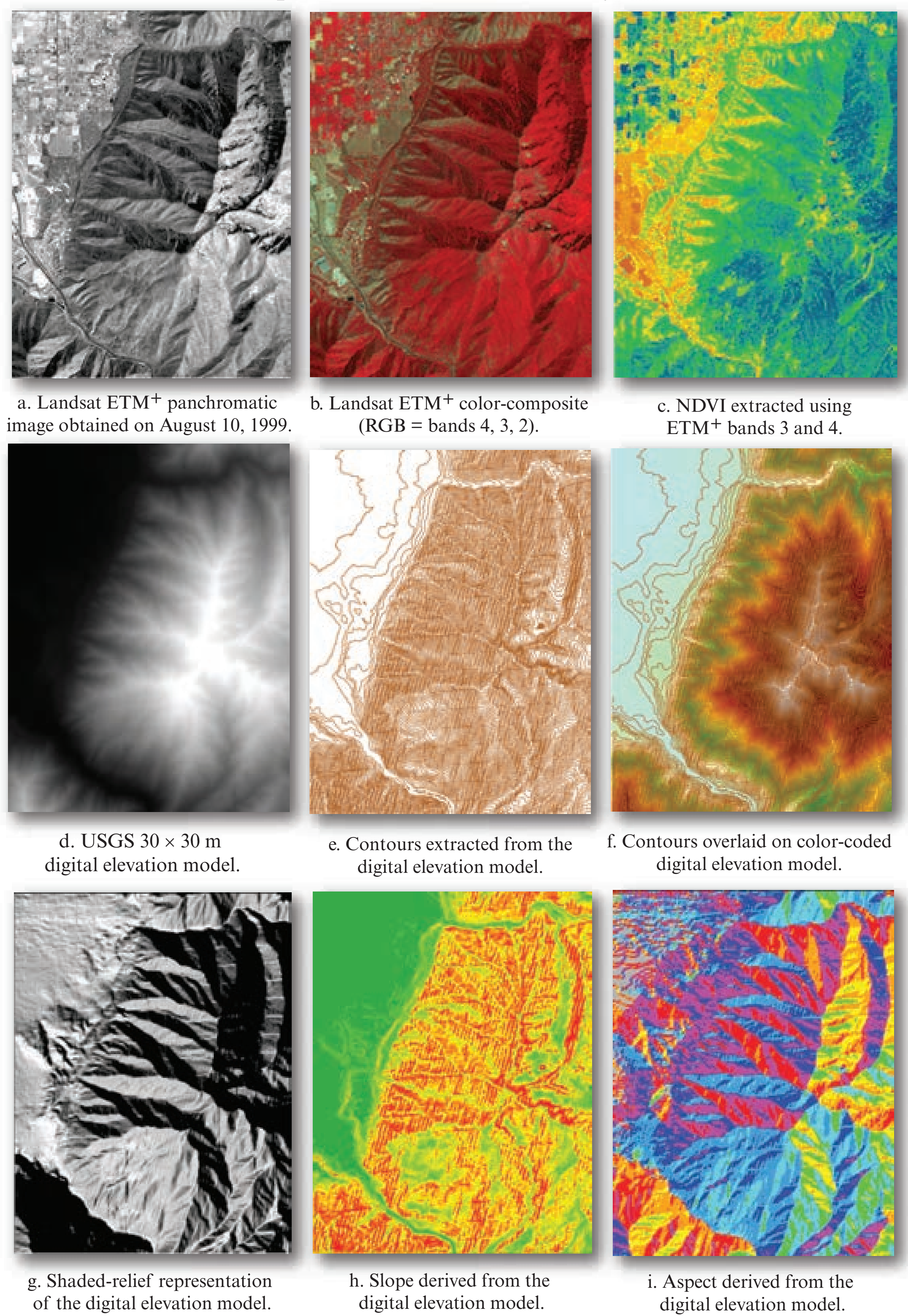


FIGURE 10–5 a–i) Datasets used to identify white fir (*Abies concolor*) on Maple Mountain in Utah County, UT (Landsat ETM$^+$ data courtesy of NASA and digital elevation model courtesy of USGS).

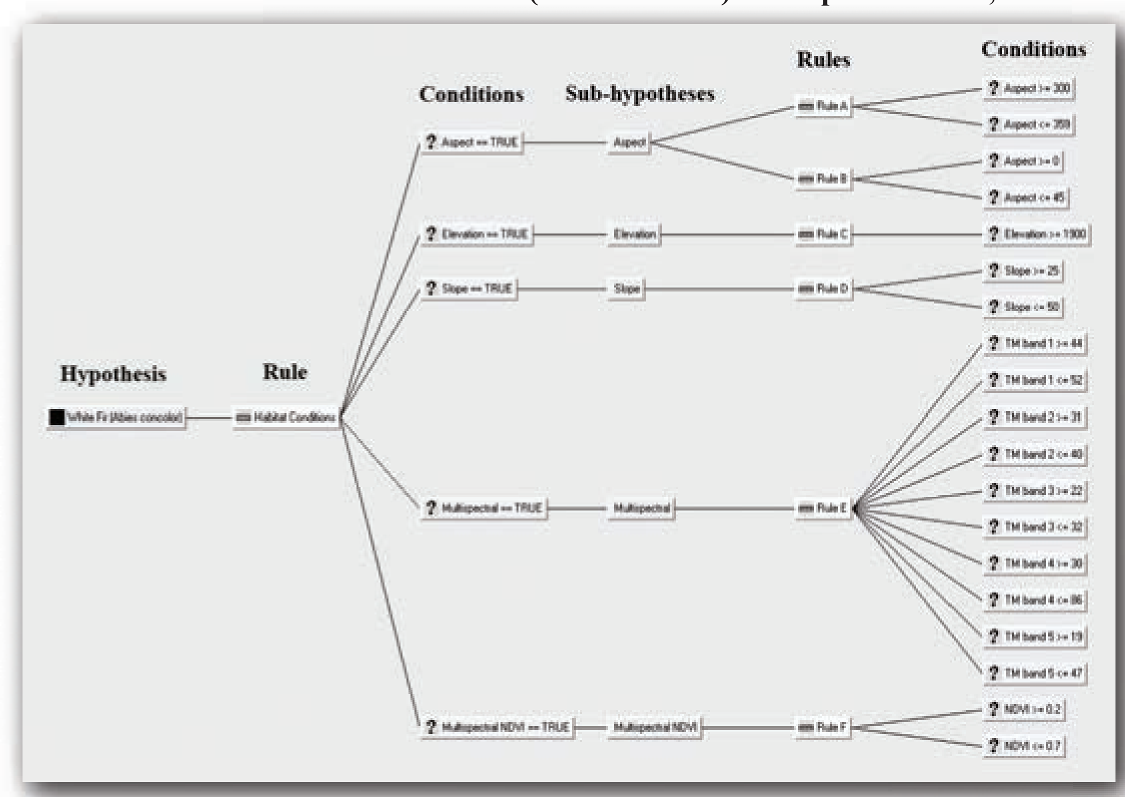


FIGURE 10–6 Hierarchical decision tree associated with a knowledge-based expert system classification of white fir (*Abies concolor*) on Maple Mountain, UT. A human expert identified the hypotheses, rules, and conditions.

previously discussed. The human domain expert explicitly expresses his or her knowledge about a subject in a language that can be understood by the knowledge engineer. The knowledge engineer translates the domain knowledge into a computer-usable format and stores it in the knowledge base.

This process presents a well-known problem in creating expert systems that is often referred to as the *knowledge acquisition bottleneck* (Jensen et al., 2009). The reasons are:

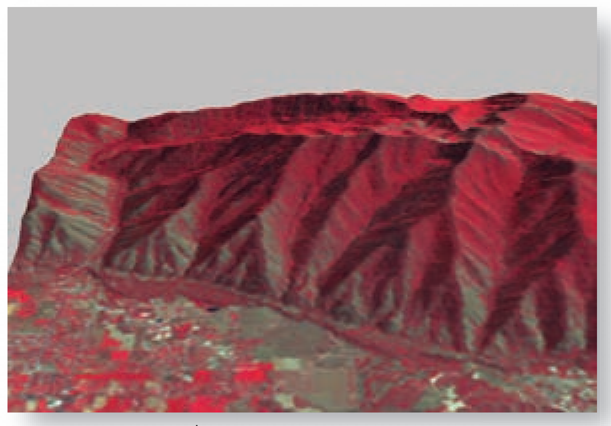
- the process requires the engagement of the domain expert and/or knowledge engineer over a long period of time, and/or
- although experts are capable of using their knowledge for decision-making, they are often incapable of articulating their knowledge explicitly in a format that is sufficiently systematic, correct, and complete to be used in a computer application.

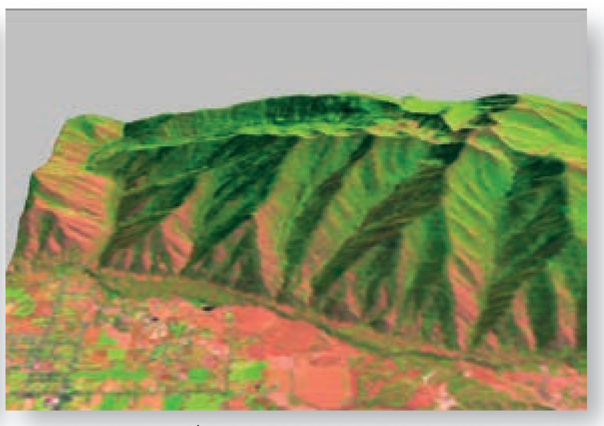
Remote sensing scientists have acknowledged these difficulties when building knowledge bases for image analysis. For example, operational monitoring of land cover from satellite requires automated procedures for analyzing large volumes of data through time (Friedl et al., 2010). To solve such problems, much effort has been exerted to automate the building of expert system knowledge bases and production rules (e.g., Huang and Jensen, 1997; Tso and Mather, 2001; Tullis and Jensen, 2003; Jensen et al., 2009; Im et al., 2012ab).

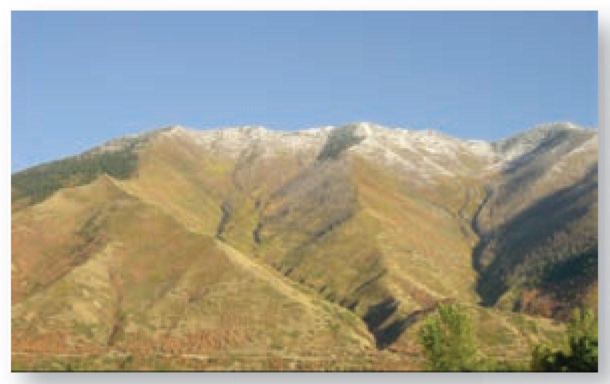
Machine Learning

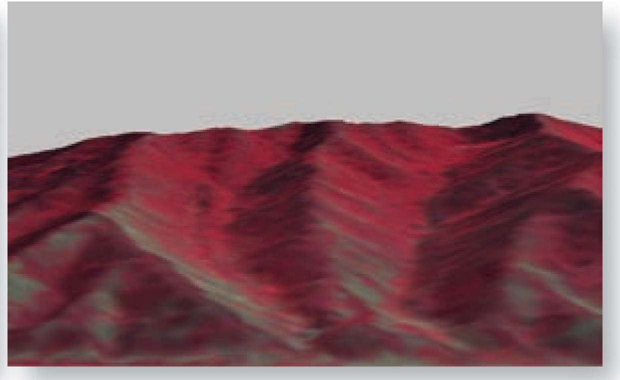
Machine learning is defined as the science of computer modeling of learning processes. It enables a computer to acquire knowledge from existing data or theories using certain inference strategies such as induction or deduction. We will focus only on inductive learning and its application in building knowledge bases for image analysis expert systems.

A human being has the ability to make accurate generalizations from a few scattered facts provided by a teacher or the environment using inductive inferences. This is called *inductive learning* (Huang and Jensen, 1997). In machine learning, the process of inductive learning can be viewed as a heuristic search through a space of symbolic descriptions for plausible general descriptions, or concepts, that explain the input training

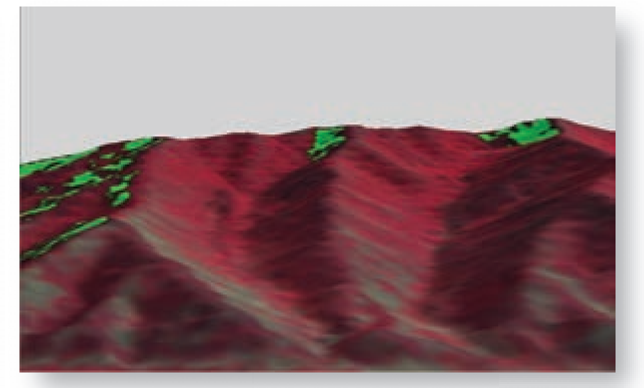

a. Landsat ETM+ (RGB = bands 4, 3, 2) image obtained on August 10, 1999, draped over a digital elevation model.


b. Landsat ETM+ (RGB = bands 5, 4, 2) image draped over a digital elevation model (azimuth = 90°).


c. Terrestrial photograph of Maple Mountain (azimuth = 113°).


d. Landsat ETM+ (RGB = bands 4, 3, 2) image draped over a DEM viewed from the same azimuth as in part c.


e. Expert system classification of white fir.


f. Expert system classification of white fir draped over the color composite and digital elevation model.

Expert System Classification of White Fir (*Abies concolor*) on Maple Mountain in Utah County, UT

FIGURE 10–7 Expert system classification of white fir (*Abies concolor*) on Maple Mountain in Utah County, UT. ab) Landsat ETM+ color composites draped over a digital elevation model. c) Terrestrial natural color photograph. d) Maple Mountain viewed from an azimuth of 113°. e,f) Vertical and terrestrial views of the white fir classification derived using human-specified rules and an expert system.

data and are useful for predicting new data. Inductive learning can be formulated using the following symbolic formulas (Michalski, 1983):

$$\forall i \in I \qquad (E_i \Rightarrow D_i) \tag{10.2}$$

$$\forall\ i, j \in I \qquad (E_i \Rightarrow \sim D_i),\ \text{if}\ (\ j\ \neq i) \tag{10.3}$$

where D_i is a symbolic description of class i, E_i is a predicate that is true only for the training events of class i, I is a set of class names, ~ stands for "negation," and $\Rightarrow$ stands for "implication." Expression 10.2 is called the *completeness condition* and states that every training event of some class must satisfy the induced description D_i of the same class. However, the opposite does not have to hold because D_i is equivalent to or more general than E_i. This means that D_i may include some features that do not exist in some examples in E_i. Expression 10.3 is called the *consistency condition* and states that if an event satisfies a description of some class, it cannot be a member of a training set of any other class. The task of inductive learning is to find through the space of descriptions the general description set $D = \{D_1, D_2, ..., D_i\}$ for the class set $K = \{K_1, K_2, ..., K_i)$ that satisfies the completeness condition and also, in some cases, the consistency condition (Huang and Jensen, 1997).

The general description set, or concept, *D,* resulting from inductive learning can be represented by a variety of formalisms, including *production rules.* This means that inductive learning can be used to build knowledge bases for expert systems because production rules are the most popular form of knowledge representation in expert systems. A motivation for the use of this approach to build a knowledge base is that it requires only a few good examples to function as *training data.* This is often much easier than explicitly extracting complete general theories from the domain expert.

It is important to remember that machine-learning *decision trees* can generate categorical output information (e.g., forest, agriculture, urban land cover) (e.g., Xian et al., 2009), while *regression trees* can predict continuous quantitative output such as leaf-area-index (LAI) or percent canopy cover (Clark Labs CTA, 2014; Xian et al., 2011; Im et al., 2012b).

There are a number of machine-learning, data-mining decision tree/regression tree programs that can be used to develop the production rules, including:

- Salford Systems (2014a), Inc. CART® Classification and Regression Trees. CART is based on mathematical theory introduced in 1984 by four statisticians at Stanford University and the University of California at Berkeley. Salford Systems' implementation of CART is the only decision-tree software embodying the original proprietary code. The CART creators continue to collaborate with Salford Systems to provide CART with proprietary advances. CART has been shown to be very effective for land-cover classification (e.g., Sexton et al., 2013).
- TIBCO S-Plus (2014) includes >4,200 data modeling and statistical analysis functions including extensive decision-tree/regression-tree creation and analysis.
- R Development Core Team (2014) provides the R language and environment for statistical computing and graphics. Tree models are available in R via the user-contributed packages *rpart* and *tree* (Therneau et al., 2014).
- RuleQuest Research, Inc. (2014) C5.0® (Unix, Linux), See5® (for Windows) and Cubist® developed by J. Ross Quinlan. C4.5 and C5.0 have been used extensively for land-cover classification and for predicting continuous variables such as leaf-area-index (e.g., Im and Jensen, 2005; Sexton et al., 2013; Im et al., 2012ab). IDRISI's Classification Tree Analysis (CTA) uses C4.5 (Clark Labs, CTA, 2014). The NASA global MODIS Collection 5 Land Cover type product (MCD12Q1) is based on the use of C4.5 (Friedl et al., 2010).

CART, *S-Plus*, *C4.5*, *C5.0*, *Cubist*, and R were not developed specifically for remote sensing digital image processing. Rather, they were developed for a diverse array of intensive data mining applications in such fields as pharmaceutical and biomedical data analysis, population studies, medical image analysis, and other data-rich applications. Fortunately, the decision rules derived using these machine-learning programs are routinely used to build a decision- and/or regression-tree classifier within remote sensing digital image processing programs such as ENVI's Interactive Decision Tree tool (Exelis ENVI, 2014), ERDAS Imagine's Knowledge Engineer (Hexagon ERDAS Imagine, 2014), or Clark Labs Classification Tree Analysis (Clark Labs CTA, 2014).

The C5.0 program will be used to demonstrate the use of machine-learning decision trees to create production rules. Most decision-tree programs have the following advantages: (Quinlan, 2003; Lu and Weng, 2007; Clark Labs CTA, 2014; Jensen et al., 2009; Exelis ENVI, 2014; RuleQuest, 2014):

- To maximize interpretability, its classifiers are expressed as decision trees or sets of if–then rules, forms that are generally easier to understand than neural networks.

- The knowledge learned can be stored in a production rule format that can be used to create a knowledge base for a rule-based expert system.
- It is flexible. Unlike many statistical approaches, it does not depend on assumptions about the distribution of attribute values or the independence of the attributes themselves. This is very important when incorporating ancillary GIS data with remotely sensed data because they usually have different attribute value distributions, and some of the attributes may be correlated.
- It is based on a decision-tree learning algorithm that is one of the most efficient forms of inductive learning. The time taken to build a decision tree increases only linearly with the size of the problem.
- *Weights* can be applied to different classes and *boosting* may take place, which reduces bias during supervised learning.

The procedure of applying the inductive learning technique to automatically build a knowledge base for a remote sensing image analysis expert system that incorporates GIS data involves *training*, *decision-tree generation*, and the creation of *production rules.* The resultant production rules compose the knowledge base and can be used by an expert system inference engine to perform the final image classification, e.g., using ENVI's interactive decision-tree tool or IDRISI's classification tree analysis.

Decision-Tree Training

A suitable classification system and a sufficient number of training samples are prerequisites for a successful training (Lu and Weng, 2007).The objective of training is to provide examples of the concepts to be learned. When building a knowledge base for image classification, the examples should be a set of training objects, each of which is represented by an attribute value class vector such as:

[attribute_1, ..., attribute_*n*, class_*i*]

Attributes are then collected for each class of information (e.g., spectral reflectance values in various bands, elevation, aspect, slope). Stratified random sampling may be the most appropriate way to locate the candidate training class pixels because it guarantees that a minimum number of samples are selected from each class (strata) (Congalton and Green, 2009). The learning algorithm attempts to extract from this attribute training dataset some generalized concepts, i.e., rules that can be used to classify the remaining data (Im et al., 2012b).

Decision-Tree Generation

The C5.0 learning algorithm first generates decision trees from the training data. The decision trees are then transformed into production rules. A raw decision tree (prior to the formation of production rules) can be viewed as a classifier composed of leaves that correspond to classes, decision nodes that correspond to attributes of the data being classified, and arcs that correspond to alternative values for these attributes. A hypothetical example of a decision tree is shown in Figures 10-8a and b.

A recursive "divide and conquer" strategy is used by C5.0 to generate a decision tree from a set of training data. The training dataset S is divided into subsets S_1, ..., S_n according to a_1, ..., a_n, which are the possible values of a single attribute A. This generates a decision tree with A being the root and S_1, ..., S_n corresponding to subtrees T_1, ..., T_n (Figure 10-8c).

The stop condition for such a procedure will eventually be satisfied, resulting in a final decision tree. The goal is to build a decision tree that is as small as possible. This ensures that the decision-making by the tree is efficient. The goal is realized by selecting the most informative attribute at each node so that it has the power to divide the dataset corresponding to the node into subsets as pure (homogeneous) as possible (DeFries and Chan, 2000). C5.0's attribute selection criteria are based on the entropy measure from communication theory where, at each node, the attribute with the minimum entropy is selected to divide the dataset.

From Decision Trees to Production Rules

Decision trees are often too complex to be understood, especially when they are large. A decision tree is also difficult to maintain and update. Therefore, it is desirable to transform a decision tree to another representation of knowledge that is adopted commonly in expert systems, such as production rules.

A *production rule* can be expressed in the following general form (Jackson, 1990):

$$P_1, ..., P_m \rightarrow Q_1, ..., Q_m \quad (10.4)$$

with the meaning

- if *premises* (or conditions) P_1 and ... P_m are true,
- then perform *actions* Q_1 and ... Q_m.

In fact, each path from the root node to a leaf in a decision tree can be translated into a production rule. For example, the path from the root node to the left most

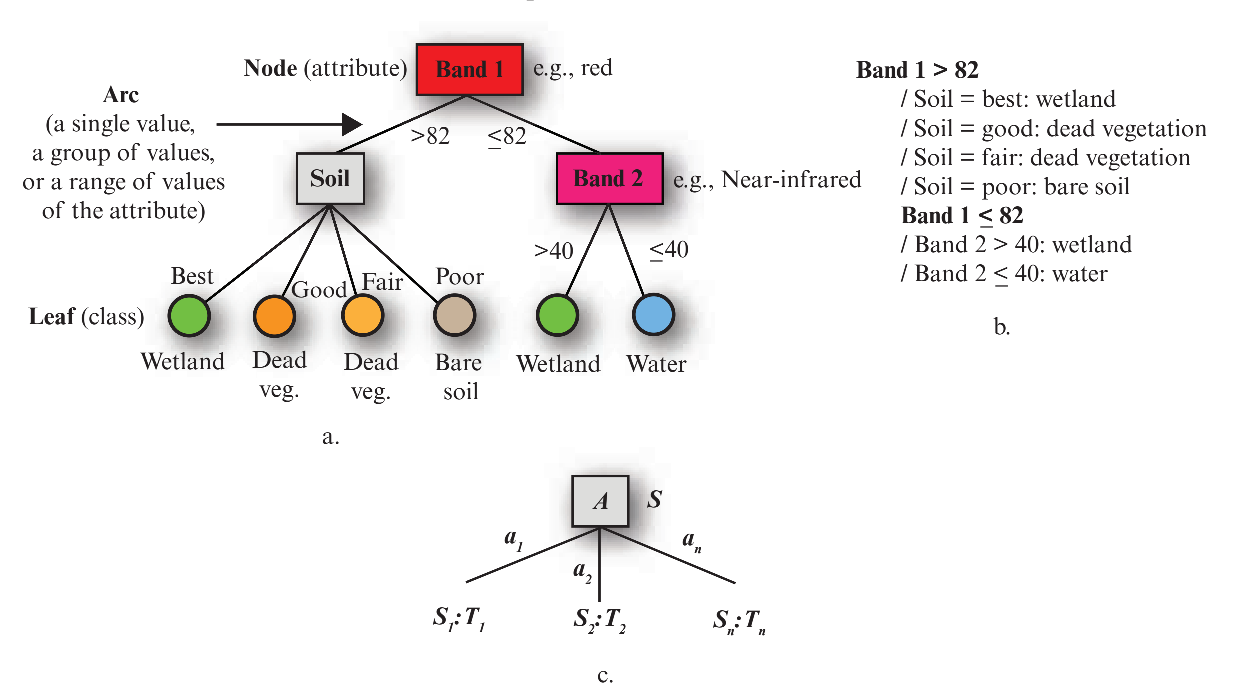


FIGURE 10–8 Example of a decision tree. a) A dataset consisting of three attributes (band 1, band 2, and soils) is classified into one of four classes: water, wetland, dead vegetation, and bare soil. b) Another method of presenting the decision tree shown in (a). c) A decision tree generated by dividing dataset *S* into subsets (based on Huang and Jensen, 1997).

leaf in the decision tree in Figure 10-8a can be represented by a production rule such as

(band 1 > 82), (soil = poor) → (class = bare soil).

Several problems must be solved when transforming a decision tree into production rules. First, individual rules transformed from the decision tree may contain irrelevant conditions. C5.0 uses a pessimistic estimate of the accuracy of a rule to decide whether a condition is irrelevant and should be deleted. Second, the rules may cease to be mutually exclusive and collectively exhaustive. Some rules may be duplicative or may conflict. This is a common problem for rule-base building using either a manual or automated approach. Usually, a rule-based system has some conflict resolution mechanism to deal with this problem. The approach adopted by C5.0 includes the use of rule confidence voting and a default class. In other words, each rule is assigned a confidence level based on false-positive errors (the number of training objects that were incorrectly classified as class C by a rule). If an object can be assigned to more than one class by two or more rules, a summary vote of all rule confidences concerned is tabulated and the class with the highest number of votes wins (Quinlan, 2014).

Unfortunately, some objects to be classified may satisfy no rules or they may satisfy conflicting rules with equal confidence levels. When this occurs, a default class might be used. There are a number of ways a default class can be identified. One approach is to have the default class be the one that contains the most training objects not satisfying any rule.

The quality of the resultant rules can be evaluated by predicting error rates derived by applying the rules on a test dataset. Because the rules are easy to understand, they can also be examined by human experts. With caution, they may be edited directly.

Case Study

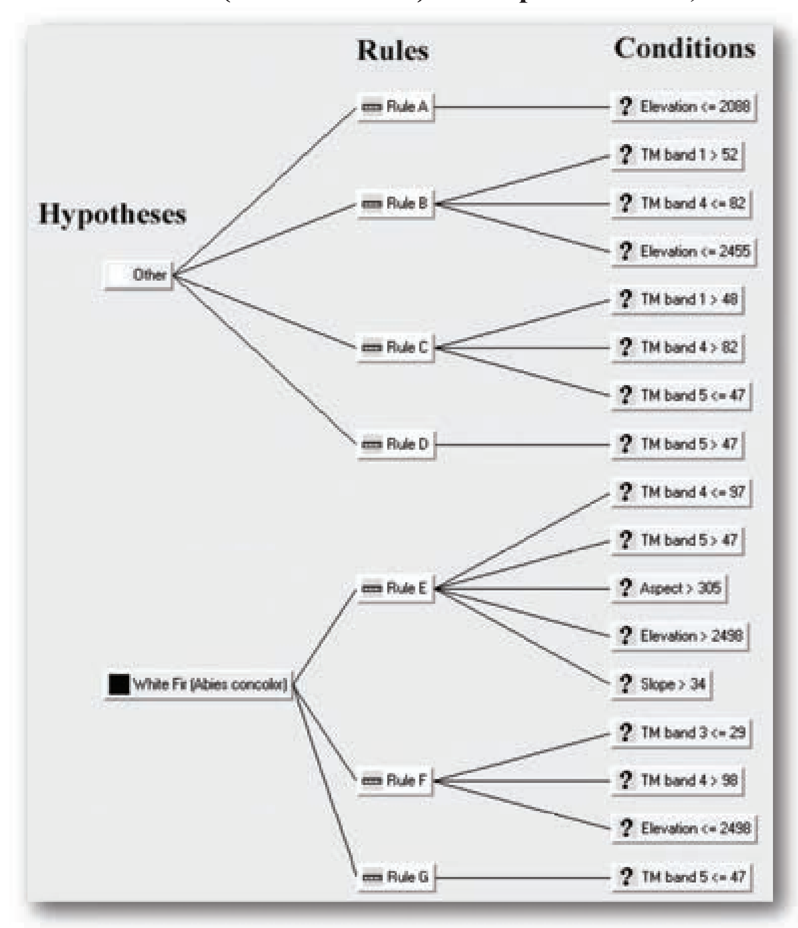
To demonstrate the utility of machine learning for the creation of production rules, the C5.0 algorithm was trained with the Maple Mountain data previously discussed. This resulted in the creation of the rules summarized in Figure 10-9a. The classification map produced from the machine-learning rules is shown in Figure 10-9b. A per-pixel confidence map of the decision-tree classification is shown in Figure 10-9c.

Advantages of Decision-Tree Classifiers

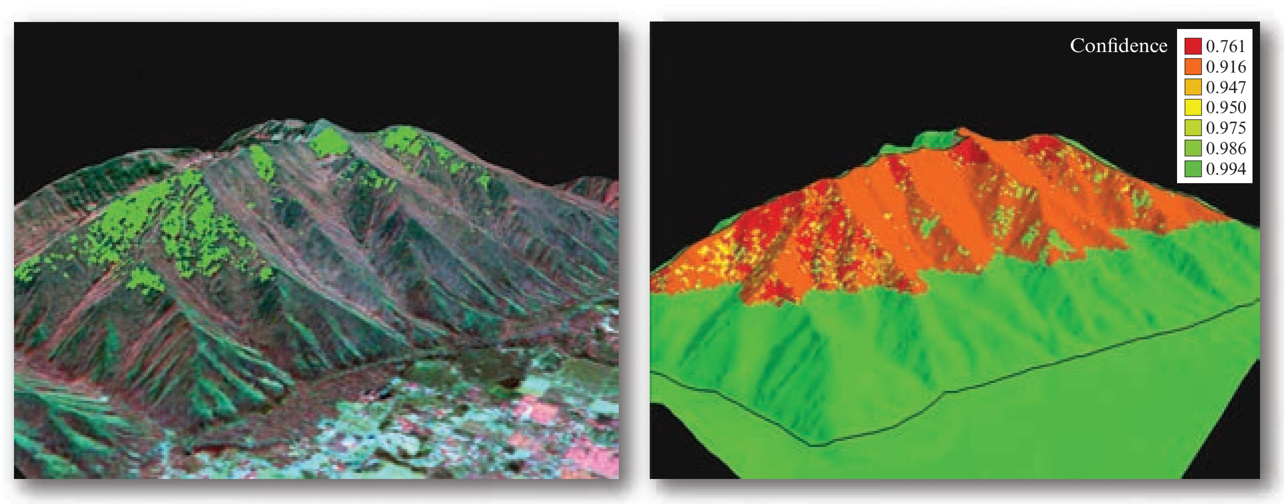
Expert system decision-tree classifiers have several characteristics that make them attractive when compared with other classification methods:

- Machine-learning algorithms may be used to train expert systems by creating rules and conditions

Machine-learning Hierarchical Decision-Tree Classification of White Fir (*Abies concolor*) on Maple Mountain, UT



a. Expert system rules for classifying white fir (*Abies concolor*) derived using machine learning.



b. Classification of white fir (*Abies concolor*) on Maple Mountain based on rules derived using machine learning draped over a three-dimensional view of Landsat ETM$^+$ imagery.

c. Confidence map.

FIGURE 10–9 a) Hierarchical decision-tree classification of white fir (*Abies concolor*) on Maple Mountain in Utah County, UT. The C5.0 machine-learning decision tree developed these production rules based on the characteristics of the training data. b) Results of applying the production rules. c) A confidence map of the classification.

directly from the training data without human intervention (e.g., Huang and Jensen, 1997; Im et al., 2012ab; RuleQuest, 2014; Salford Systems, 2014a). This is a very important advancement because the expert system can adapt when new learning data are provided. Otherwise, an expert system cannot learn by example.

- You can evaluate the output of the expert system (e.g., whether a certain pixel is ideal or not for residential development) and work backward to identify how a conclusion was reached. This is in contrast to neural networks where the exact nature of the decision process is often lost in the weights used in the hidden layers (Qiu and Jensen, 2004).
- Until recently, maximum likelihood classification was the most common method used for supervised classification of remote sensor data (McIver and Friedl, 2002; Lu and Weng, 2007). This method assumes that the probability distributions for the input classes possess a multivariate normal distribution. Increasingly, nonparametric classification algorithms are being used such as decision trees, which make no assumptions regarding the distribution of the data (Friedl et al., 2010).
- The decision tree can reveal nonlinear and hierarchical relationships among the input variables and use them to predict class membership.

A large body of evidence demonstrates the ability of machine-learning techniques (particularly decision trees and neural networks) to deal effectively with tasks that involve multi-dimensional data (e.g., hyperspectral). Reducing a dataset to just two or three variables (e.g., bands) is becoming an outdated notion (Gahegan, 2003). The new thinking is to let the geographical data themselves "have a stronger voice" rather than let statistics derived from the dataset dictate the analysis (e.g., the means and covariance matrices used in maximum likelihood classification).

Random Forest Classifier

Now that we are comfortable with single decision-tree classifiers, it is important to note that there are also machine-learning algorithms that make use of multiple decision trees (i.e., an ensemble) at one time during the classification process. *Ensemble* learning algorithms (e.g., Random Forests, bagging, boosting) have emerged as potentially more accurate and robust alternatives to conventional parametric (e.g., maximum likelihood) or other machine-learning algorithms (e.g., individual decision trees, neural networks) (Rodriguez-Galiano et al., 2012ab; Chen et al., 2014).

Random Forests grows many classification decision trees. To assign an unclassified pixel (or OBIA object) with its associated attributes (e.g., spectral reflectance, elevation, slope) into a class, the unclassified input pixel values are run through each of x decision trees in the forest (e.g., 200 trees). Each decision tree classifies the pixel into one of y classes, i.e. it "votes" for that class. The forest assigns the pixel to the class having the most votes from all the trees in the forest (Breiman, 2001; Breiman and Cutler, 2014; Hayes et al., 2014).

Each of the individual decision trees is grown as follows (Breiman and Cutler, 2014):

1. If the number of cases in the training set is N, sample N cases at random, but *with* replacement, from the original data. This sample will be the training set for growing the tree. Generally, for every tree grown, only about 70% of the training data is used and about 30% of the training data are left out-of-bag (OOB), i.e., they are not used.
2. If there are M input variables, a number $m < M$ is specified such that at each node, m variables are selected at random out of the M and the best split of these m is used to split the node. The value of m is held constant as the forest of trees is grown. It is suggested that m be the square root of M.
3. Each decision tree is grown to the largest extent possible. There is no pruning.

Individual classes can be weighted *a priori* if desired. Random Forests runs efficiently on very large datasets, which is a useful characteristic when analyzing remote sensor data. It also provides estimates of what variables are most important in the classification.

Numerous studies have evaluated Random Forest classification with other types of image classification (e.g., Pal, 2005; Guo et al., 2011; Rodriguez-Galiano et al., 2012ab). Random Forest freeware is available in R via the user-contributed package *randomForest* (R Development Core Team, 2014; Liaw and Weiner, 2014). Random Forest commercial software is available through Salford Systems (2014b).



Support Vector Machines

Support vector machines (SVMs) are another type of machine-learning software that can be used for remote sensing classification (Pal and Mather, 2005; Jensen et al., 2009; Duro et al., 2012). Comparative studies of SVMs versus decision trees and artificial neural networks have demonstrated the robustness of SVMs, especially for small training sets (e.g., Huang et al., 2002;

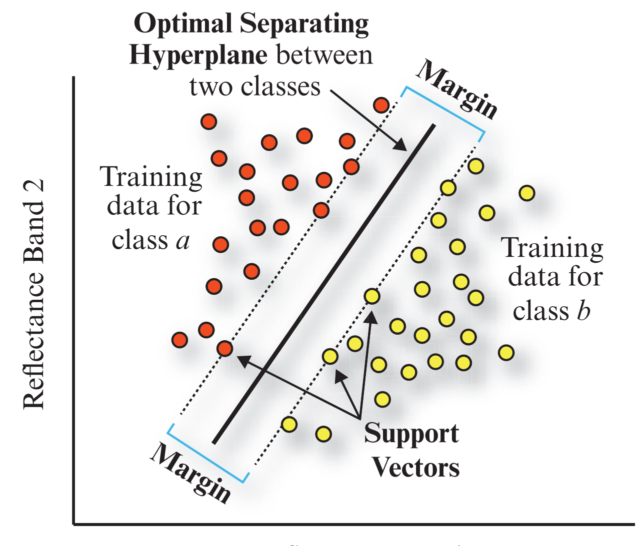


FIGURE 10–10 A support vector machine classification relies heavily on the training class samples on the edge of the margin to identify the location of the separating hyperplane (based on Meyer, 2014).

Foody and Mathur, 2004). SVMs find an optimal separating hyperplane (OSH) between classes using the training samples (Foody and Mathur, 2006; Van der Linden and Hostert, 2009). However, in this case we are looking for the optimal separating hyperplane between two classes by maximizing the margin between the classes' closest training samples, as shown in Figure 10-10. The points lying on the boundaries are called *support vectors*, and the middle of the margin is the optimal separating hyperplane (Meyer, 2014). Training data points on the "wrong" side of the discriminant margin are given negative weights to reduce their influence. When it is not possible to find a linear separator such as shown in Figure 10-10, the data points can be projected into a higher-dimensional space (using kernel techniques). A program that performs all of these tasks is called a **support vector machine** (Meyer, 2014).

SVMs focus on solving the optimization problem with training samples using the following equations (Jensen et al., 2009):

$$min_{w,b,\zeta}\left(\frac{w^T \cdot w}{2} + C\sum_{i=1}^{\lambda}\zeta_i\right) \qquad (10.5)$$

$$y_i(w^T \cdot \Phi(x_i) + b) \geq 1 - \zeta_i, \qquad (\zeta_i \geq 0)$$

where ζ_i are positive slack variables used to allow some of the samples to fall on the wrong side of the hyperplane, $C\,\Sigma\,\zeta_i$ is a term used to penalize solutions for which ζ_i are very large, and $w^T \cdot \Phi(x_i) + b$ is a hyperplane in a highly dimensional feature space (Su et al., 2006). The basic SVM approach may be extended to optimize nonlinear surfaces using the following decision function (Jensen et al., 2009):

$$f(x) = \sum_{i=1}^{\lambda} a_i y_i K(x, x_i) + b \qquad (10.6)$$

where a_i are nonnegative Lagrange multipliers used to search the OSH, and $K(x, x_i)$ is a kernel function, which replaces the inner product $(x \cdot x_i)$ in order to solve computational problems in a higher dimensional space (Wang et al., 2005). The selection of the kernel function k is critical for producing successful results. Several widely used SVM types of kernels include linear, polynomial, Gaussian, sigmoid, and spectral angle mapper kernels.

One advantage of SVMs is that small training samples selected on the boundaries of classes, where mixed pixels are common, can lead to accurate classification (Foody and Mathur, 2006). Pal (2005) compared the accuracies of two supervised classification algorithms using Landsat ETM$^+$ data: Random Forests and SVMs and found that they performed equally well (Duro et al., 2012). SVM freeware is available in R via the user-contributed package *e1071* (R Development Core Team, 2014; Meyer, 2013, 2014).



Neural Networks

A neuron is a cell in the brain whose principal function is the collection, processing, and dissemination of electrical signals (Russell and Norvig, 2010). Neural networks simulate the thinking process of human beings, whose brains are a large number of small and interconnected processors or neurons that work together to solve difficult classification and optimization problems (Khorram et al., 2011). A neural network reaches a solution not via a step-by-step algorithm or a complex logical program, but in a non-algorithmic, unstructured fashion based on the adjustment of the weights connecting the neurons in the network (Filippi et al., 2010). Neural networks have been used to classify various types of remote sensor data and have in certain instances produced results superior to those of traditional statistical classification methods (e.g., Foody, 1996; Jensen et al., 1999; Ji, 2000; Aitkenhead and Aalders, 2011; Santi et al., 2014). This success can be attributed to two of the important advantages of neural networks: 1) freedom from normal distribution requirements (Hu and Weng, 2009; Khorram et al., 2011), and 2) the ability to adaptively simulate complex

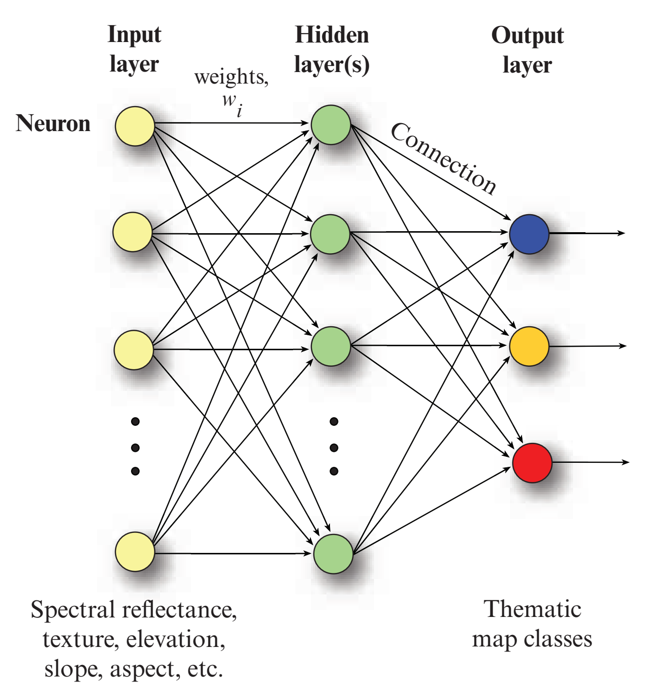


FIGURE 10–11 Components of a typical back-propagation artificial neural network (ANN). The network consists of input, hidden, and output layers. The input layer may contain information about individual training pixels including spectral reflectance in various bands and ancillary data such as elevation, slope, etc. Each layer consists of nodes that are interconnected. This interconnectedness allows information to flow in multiple directions (i.e., back-propagation can take place) as the network is trained. The strength (or weight) of these interconnections is eventually *learned* by the neural network and stored. These weights are used during the *testing* (classification) procedure. The more representative the training data, the more likely that the neural network will develop weights in the hidden layer that mirror reality and result in an accurate classification. The output layer may represent individual thematic map classes such as water or forest (based on Jensen, J. R., Qiu, F. and K. Patterson, 2001, "A Neural Network Image Interpretation System to Extract Rural and Urban Land Use and Land Cover Information from Remote Sensor Data," *Geocarto International*, 16(1):19–28).

and nonlinear patterns given proper topological structures (Filippi and Jensen, 2006).

Components and Characteristics of a Typical Artificial Neural Network Used to Extract Information from Remotely Sensed Data

The topological structure of a typical back-propagation neural network is shown in Figure 10-11. The artificial neural network normally contains neurons arranged in three types of layer:

- an input layer,
- a hidden layer(s), and
- an output layer.

The neurons in the input layer might be the multispectral reflectance values associated with individual pixels in the imagery and/or objects if object-based image analysis is (OBIA) is used. However, one of the main advantages of neural networks is the fact that multiple data sources, including non-spectral data, can be used as input such as elevation, aspect, slope, etc. (Brown et al., 2008).

The use of neurons in the hidden layer(s) enables the simulation of nonlinear patterns in the input data. The choice of how many hidden layers and the number of nodes per hidden layer is an ongoing basic research question (Brown et al., 2008; Hu and Weng, 2009). Aitkenhead and Aalders (2011) suggest that each hidden layer used should have twice as many nodes as the number of nodes in the input layer based on Kolmogorov's theorem for neural network modelling of continuous functions.

Land-cover classes can be mapped using an output layer that contains a neuron for each class. Or, it is possible to use only a single output neuron to predict continuous variables such as wheat yield or chlorophyll *a* concentration as demonstrated by Jiang et al. (2004) and Gonzalez Vilas et al. (2011).

Much like a supervised classification, an artificial neural network requires *training* and *testing (classification)* to extract useful information from the remotely sensed and ancillary data (Atkinson and Tatnall, 1997; Foody and Arora, 1997; Qiu and Jensen, 2004). ANN model development can also take place using unsupervised techniques (Filippi et al., 2010).

Training an Artificial Neural Network

In the *training* phase, the analyst selects specific *x, y* locations in the input image with known attributes (e.g., agriculture, upland pine) as training sites (Jensen et al., 2001; Li and Eastman, 2006; Li, 2008). The per-pixel spectral information (e.g., spectral reflectance in red and near-infrared bands) and ancillary information (e.g., elevation, slope, aspect) for each training site is then collected and passed to the input layer of the neural network. At the same time, the true target (class) value of this exact location (e.g., agriculture) is sent to the output layer by assigning the neuron representing

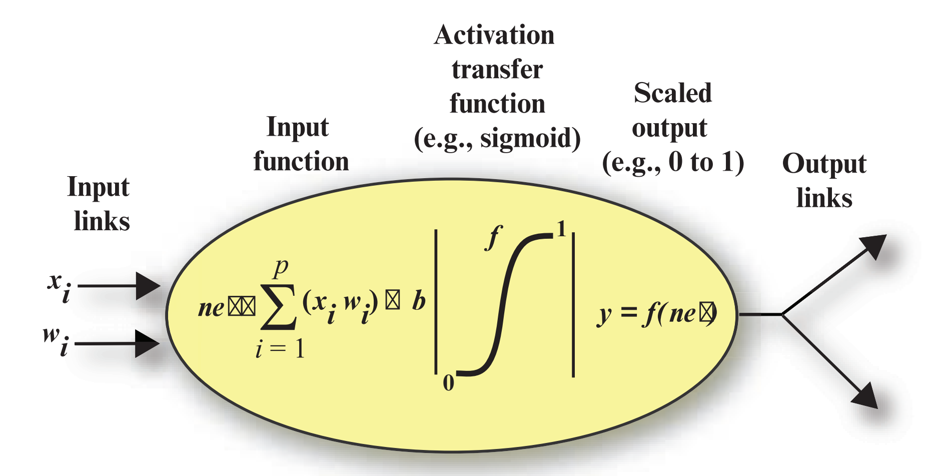


FIGURE 10–12 Mathematical model of a neuron. The output of the neuron is y, which is a function of the summation of the input values (x_i) multiplied times their individual weights (w_i) in conjunction with an activation function (f). A sigmoid activation function is used in this example.

this class a membership value of 1, while all the other output neurons are assigned a value of 0.

It is important to remember that neural network training based on the examples obtained from an image and other ancillary data acquired at a specific time and location may be applicable only in the immediate geographic area and perhaps only for a given season. Thus, they may not be extendable through space or time.

Learning is usually accomplished by adjusting the weights using a back-propagation algorithm (Brown et al., 2008). For each training example, the output of the network is compared with the true target (class) value. The difference between the target and output value is regarded as an error and is passed back to previous layers in the network to update the connection weights. The magnitude of the adjustment is proportional to the absolute values of the error (discussed shortly in the mathematics section). After many iterations, the root mean square (RMS) error diminishes to a small value less than a predefined tolerance, and further iterations will not improve the performance of the network (Jensen et al., 1999). At this time, the system achieves convergence and the training process is completed. The rules inherent in the examples are stored in the hidden weights in the hidden layer(s) for use in the testing (classification) phase.

Testing (Classification)

During the *test* or *classification* phase, the spectral and ancillary characteristics of every pixel in the scene are passed to input neurons of the neural network. The neural network evaluates each pixel using the weights stored in the hidden layer neurons to produce a predicted value for every neuron of the output layers. The value obtained for every output neuron is a number between 0 and 1 that gives the fuzzy membership grade of the pixel belonging to the class represented by that neuron. Defuzzification of these maps using a local maximum function leads to a hard classification map where each pixel is assigned to a unique class that has the highest fuzzy membership grade (Jensen et al., 2001).

Mathematics of the Artificial Neural Network

An artificial neural network (ANN) is defined by neurons, topological structure, and learning rules. The neuron is the fundamental processing unit of an ANN for computation. Analogous to the human brain's biological neuron, an artificial neuron is composed of inputs (dendrites), weights (synapses), processing units (cell bodies), and outputs (axons) (Hagan et al., 1996). Each input x_i is multiplied by the scalar weight w_i to form $x_i w_i$, a term that is sent to the "summing unit" of the processing unit (Figure 10-12). An offset, b, may be added to the total. The summation output:

$$net = \sum_{i=1}^{p}(x_i w_i) + b , \qquad (10.7)$$

referred to as net input, goes into an activation transfer function f that produces scaled neuron output y (between 0 and 1, or –1 to 1) through a transform algorithm. Thus, y can be calculated as (Jensen et al., 1999):

$$y = f(net) = f\left[\sum_{i=1}^{p}(x_i w_i) + b\right] . \qquad (10.8)$$

The topological structure of an artificial neural network defines the overall architecture of the network,

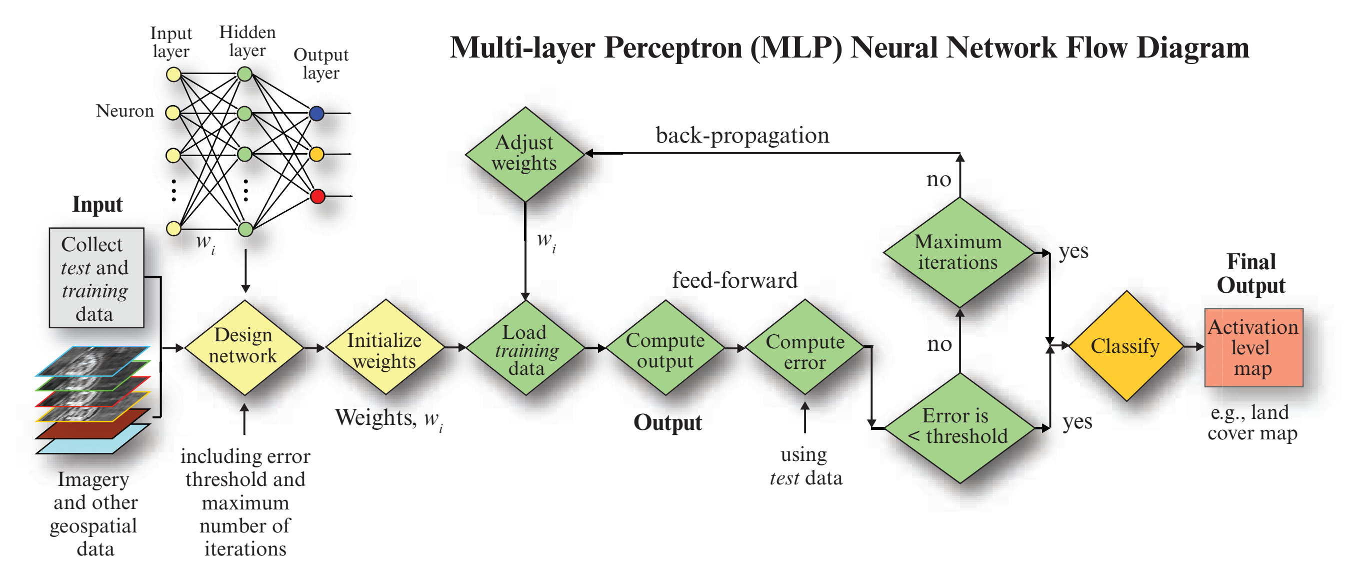


FIGURE 10–13 The flow of data (e.g., test, training, imagery), decisions, and iterations associated with a typical back-propagation multi-layer perceptron (MLP) neural network (based on Hu, X. and Q. Weng, 2009, "Estimating Impervious Surfaces from Medium Spatial Resolution Imagery Using the Self-organizing Map and Multi-Layer Perceptron Neural Networks," *Remote Sensing of Environment*, 113:2089–2102).

including the framework and interconnection of neurons that are organized into layers. As previously discussed, a typical ANN consists of three or more layers (Figure 10-11): one input layer, one output layer, and one or more hidden layers (Jensen et al., 2001). Neurons within and between layers can be connected to form a specific user-defined, task-oriented network. Decision processes can be duplicated or approximated through an ANN's ability to learn and recall. Learning is made possible by feeding the input layer with training data. By comparing the current activation of neurons in the output layer to a desired output response, the difference can be obtained and used to adjust weights connecting neurons in the network. It is the weights that determine the behavior of the network, so the goal of learning is to achieve a set of weights that will produce an output most resembling the target. This adaptive learning process repeats until changes in network weights drop below a preset threshold, which indicates a user-defined accuracy. Once learned, the network can recall stored knowledge to perform either classification or prediction on new input data.

The desirable properties of non-normality and nonlinearity of an ANN can be attributed not only to the network's massively parallel distributed structure, but also to the transfer function for each neuron (Haykin, 1994). A neuron is basically a nonlinear device that may take continuous and differential functions as the transfer function. Therefore, ANNs are capable of modeling any complex system, be it physical or human. Image classification using ANNs takes advantage of this relationship to incorporate non-normally distributed numerical and categorical GIS data and image spatial information (e.g., texture and context) into the process (Qiu and Jensen, 2004).

There are several commonly used artificial neural network learning algorithms that can be used to analyze remote sensor data, including:

- Feed Forward Multi-layer Perceptron (MLP) Neural Network with Back Propagation (BP),
- Kohonen's Self-Organizing Map (SOM) Neural Network, and
- Fuzzy ARTMAP.

It is useful to briefly introduce their characteristics.

Feed Forward Multi-Layer Perceptron (MLP) Neural Network with Back Propagation (BP)

One of the most popular neural networks used in remote sensing is the feed forward multi-layer perceptron (MLP) neural network trained via back-propagation (BP) (Filippi and Jensen, 2006; Weng and Hu, 2008; Hu and Weng, 2009). MLP neural networks typically consist of one input layer, one or more hidden layer, and one output layer (Gonzalez Vilas et al., 2011; Weng, 2012). Each input neuron represents one band of imagery (and/or other type of geospatial data), and the number of output nodes corresponds to the number of information classes desired. MLP networks using BP training algorithms are typically used for complex image classification. Classification accuracy is very sensitive to the quality and quantity of training data selected for each input class (Atkinson and Tatnall, 1997; Khorram et al., 2011).

A typical MLP neural network with a BP learning algorithm uses the following algorithm to calculate the input that a single node, j, receives (Weng and Hu, 2008):

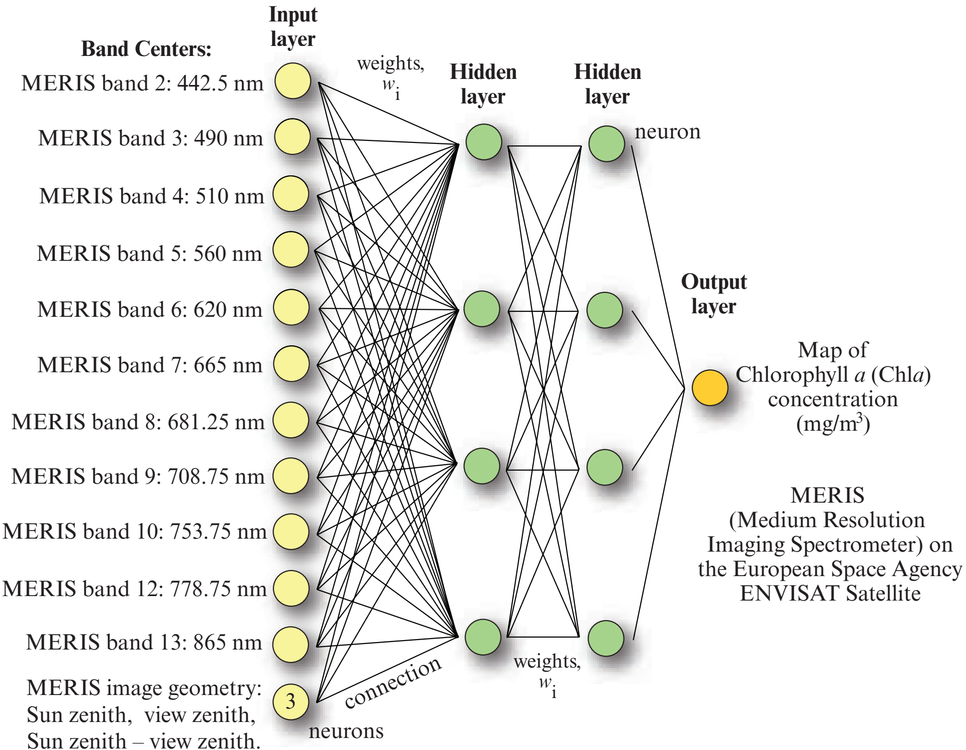


FIGURE 10–14 Structure of a multi-layer perceptron (MLP) neural network to predict chlorophyll *a* concentration of Galician *Rias* in NW Spain using 11 bands of MERIS multispectral data ranging from 442.5 to 865 nm and three geometric characteristics associated with the images including Sun zenith, view zenith, and the difference between the Sun zenith and view zenith (based on Figure 4 from Gonzalez Vilas, L., Spyrakos, E. and J. M. Torres Palenzuela, 2011, "Neural Network Estimation of Chlorophyll *a* from MERIS Full Resolution Data for the Coastal Waters of Galician *rias* (NW Spain)," *Remote Sensing of Environment*, 115:524–535).

$$net_j = \sum_i w_{ij} I_i \qquad (10.9)$$

where w_{ij} denotes the weights between nodes i and j, and I_i is the output from node i of a sender layer (either input or hidden layer). The output from a node j is calculated as (Hu and Weng, 2009):

$$O_j = f(net_j)\,. \qquad (10.10)$$

The function f is usually the nonlinear sigmoidal function previously discussed.The flow of information, decisions, and interactions associated with a typical MLP neural network are shown diagrammatically in Figure 10-13 (Hu and Weng, 2009). Note the adjustment of the weights that takes place during back-propagation until the error threshold or a user-specified maximum number of iterations is reached. The knowledge inherent in the training data is obtained through the iteration of the forward and backward passes until all the parameters in the network are stabilized (Jensen et al., 2001). This is signified by the decrease of average squared error to a minimum or acceptable level. The outputs of an MLP are called *activation level maps* with the number of activation level maps equal to the number of output nodes (Hu and Weng, 2009). The value of each pixel in the activation level map represents the degree to which a pixel belongs to a particular class.

Gonzalez Vilas et al. (2011) used a MLP ANN to map chlorophyll *a* concentration in the coastal waters of NW Spain. Their neural network consisted of 1) an input layer containing 11 MERIS multispectral bands and three geometric variables associated with the multispectral data (Sun zenith, view zenith, and the difference between Sun and view zenith), 2) two hidden layers with four nodes each, and 3) a single output neuron to map chlorophyll *a* concentration as shown in Figure 10-14.

Jiang et al. (2004) used a back-propagation ANN to predict winter wheat yield in northern China. Five input nodes were used and one hidden layer with eight nodes (Figure 10-15). Note that four of the input layer nodes are NOAA AVHRR remote sensing–derived measurements including the Normalized Difference Vegetation Index (NDVI), Absorbed Photosynthetically Active Radiation (APAR), canopy surface temperature (T_s), and a Water Stress Index. The fifth input

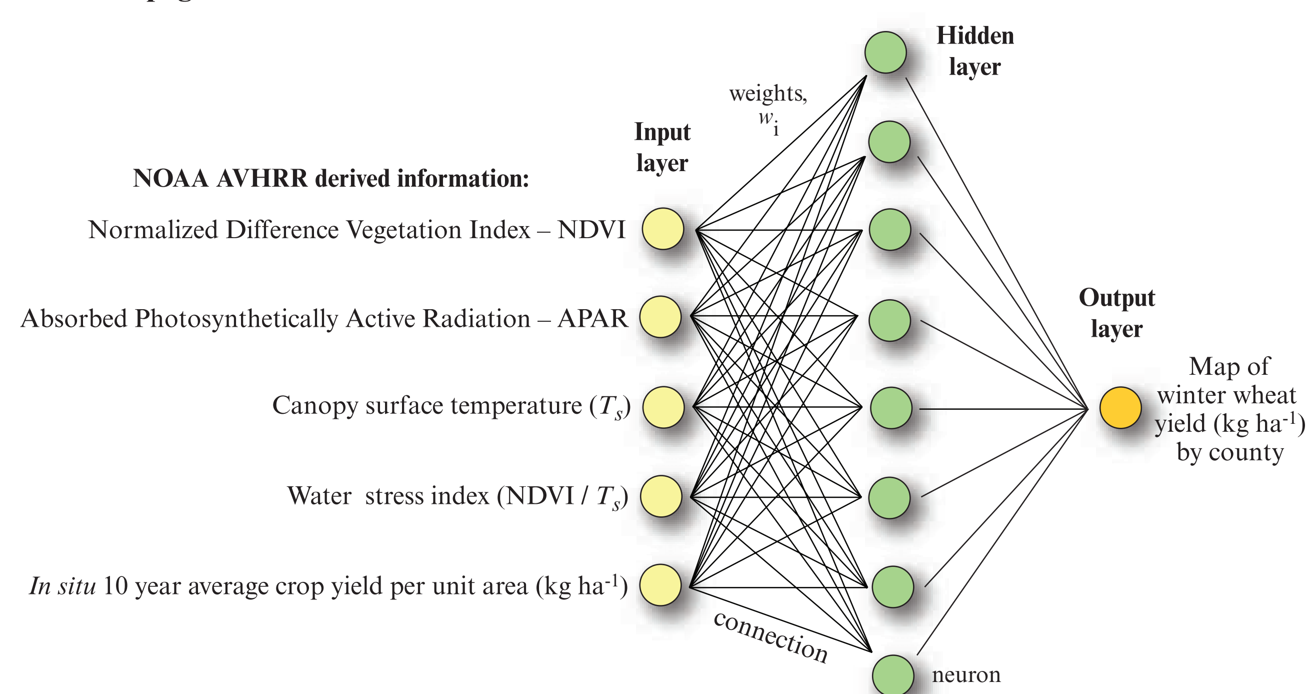


FIGURE 10–15 The characteristics of a back-propagation neural network used to predict winter wheat yield (kg ha^{-1}) in the northern provinces of China using four NOAA AVHRR-derived variables plus the average *in situ* crop yield per county over the last 10 years. Eight hidden layer nodes and a single output layer were used to map the spatial distribution of the winter wheat yield (based on Jiang, D., Yang, X., Clinton, N., and N. Wang, 2004, "An Artificial Neural Network Model for Estimating Crop Yields using Remotely Sensed Information," *International Journal of Remote Sensing*, 25(9):1723–1732).

variable (node) is average *in situ* crop yield over the last 10 years (by county). A single output layer node was used to predict winter wheat yield.

Kohonen's Self-Organizing Map (SOM) Neural Network

Kohonen's self-organizing map (SOM) is another type of neural network used in remote sensing research (Filippi et al., 2010). A SOM neural network has only one input layer and one output layer (Filippi and Jensen, 2006; Li, 2008). The output layer is known as the Kohonen or competitive layer (Li and Eastman, 2006; Filippi et al., 2010). The input layer contains neurons for each measurement dimension (e.g., image bands or other types of geospatial data such as elevation, slope, and aspect), and the output layer is usually organized as a two-dimension array or lattice of neurons, e.g., 3 × 3, 4 × 4 as shown in Figure 10-16 (Hu and Weng, 2009; Salah et al., 2009; Weng, 2012). Each output layer neuron is connected to all of the neurons in the input layer by synaptic weights (Li and Eastman, 2006). Because it only uses two layers, the dilemma of determining the hidden layer size is avoided.

Self-organization involves adaptively modifying the synaptic weights as a result of input excitations while abiding by a learning rule until a useful configuration is produced. In the output layer or lattice shown in Figure 10-16, output neurons become selectively tuned to the presented input patterns during a competitive-learning procedure. The topologically-ordered output space entails neurons close to one another representing similar input patterns. Learning is accomplished by first choosing an output neuron that most closely matches the presented input pattern, then determining a neighborhood of excited neurons around the *winner*, and finally, updating all of the excited neurons (Li, 2008; Filippi et al., 2010).

During the labeling stage of an SOM commitment (SOM-C) neural network analysis, a competitive layer neuron in the $n \times n$ output layer is triggered by different patterns. The degree of commitment indicts how much an input pattern belongs to a class and is measured using the equation (Hu and Weng, 2009):

$$C_i = \frac{P_i(j)}{\sum_{i=1}^{m} P_i(j)} \qquad (10.11)$$

Kohonen's Self-Organizing Map (SOM) Artificial Neural Network

FIGURE 10–16 Kohonen's Self-Organizing Map (SOM) neural network has one input layer and one competitive output layer consisting of an array (lattice) of neurons (e.g., 4 × 4; 5 × 5). Each output layer neuron is connected to all of the neurons in the input layer by synaptic weights (w_i). Because it uses only two layers, the dilemma of determining the hidden layer size is avoided (based on Salah et al., 2009; Hu and Weng, 2009).

where $P_i(j)$ is the proportion of training site of class i, triggering neuron j, with $P_i(j)$ calculated as

$$P_i(j) = \frac{f_i(j)}{N_i} \tag{10.12}$$

where $f_i(j)$ is the frequency of neuron j triggered by pixels labeled as class i, and N_i is the total number of samples of class i in the training sites.

Hu and Weng (2009) compared MLP with SOM to estimate impervious surfaces at the sub-pixel level using three ASTER images of Marion County, IN, and found that the SOM neural network outperformed the MLP neural network slightly for each season of image data, especially in residential areas.

Fuzzy ARTMAP Neural Network

Adaptive Resonance Theory (ART) based neural networks evolved from the biological theory of cognitive information processing (Mannan et al., 1998). A generalized fuzzy version was introduced by Carpenter et al. (1997). The fuzzy ARTMAP architecture designed for supervised classification consists of two fuzzy ART modules, ART_a and ART_b as shown in Figure 10-17. The ART_a module is composed of two layers of processing elements known as F_1 (input signals) and F_2 (output categories). The F_1 layer contains the input remote sensing or other type of geospatial information used to train the network. The number of neurons in the F_2 layer is dynamically determined during the process of learning (Mannan et al., 1998) (Figure 10-17). A controller uses a minimum learning rule to minimize predictive error and maximize predictive generalization (Carpenter et al., 1997). This enables the system to learn and to determine the number of hidden units needed to meet the matching standards (Rogan et al., 2008; Filippi et al., 2009).

The ART_b module consists of a single Output layer that has n neurons associated with the n classes of interest. There are seven classes in the example shown in Figure 10-17. The ART_a and ART_b modules are connected by the Map Field layer, which has the same number of neurons (n) as the Output layer. The Map Field and Output layers are connected directly. ARTMAP has been shown to minimize the three main drawbacks of back-propagation networks, including: sensitivity to the choice of network parameters, overfitting, and the need for user intervention during the training phase (Mannan et al., 1998).

Mannan et al. (1998) found that fuzzy ARTMAP was more efficient than MLP with the back propagation learning in supervised classification of multispectral remotely sensed images. Rogan et al. (2008) compared fuzzy ARTMAP with two decision-tree classification programs (S-Plus and C4.5) for classification of land cover in southern California. They found ARTMAP to be robust and accurate for automated, large area change monitoring with minimal human intervention required in the classification process. Li (2008) put forth non-parametric measures associated with fuzzy ARTMAP to handle spatial uncertainty in remote sensing classification.

Advantages of Artificial Neural Networks

Characteristics of selected artificial neural networks are listed in Table 10-3. Artificial neural networks have

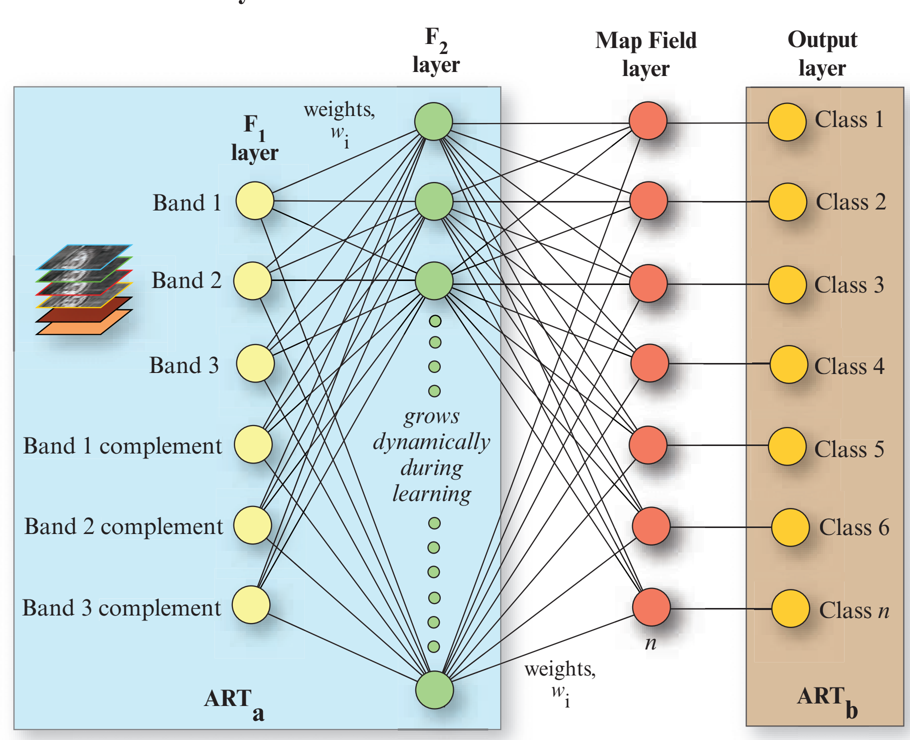


FIGURE 10–17 The architecture of a typical fuzzy ARTMAP neural network consisting of two parts, ART_a and ART_b. In this example, there are three input sources of geospatial information in the F_1 Input layer (and their complement). The number of neurons in the F_2 layer grows dynamically during the learning process. The number of Map Field layers is the same as the desired number of Output layers; in this case 7 (based on Figure 1 from Li, Z., 2008, "Fuzzy ARTMAP Based Neurocomputational Spatial Uncertainty Measures," *Photogrammetric Engineering & Remote Sensing*, 74(12):1573–1584).

been employed to process multispectral remote sensing images and often achieve improved accuracy compared to those obtained using traditional statistical methods (Filippi and Jensen, 2006; Hu and Weng, 2009; Filippi et al., 2010; Khorram et al., 2011). These successes are underpinned by many salient characteristics of neural networks, including:

- A neural network makes no *a priori* assumptions about normal and linear data distribution due to its operation in a nonparametric fashion (Rogan et al., 2008; Weng and Hu, 2008; Filippi et al., 2010).
- Neural networks are able to learn from existing examples adaptively, which makes the classification objective.
- The nonlinear patterns are "learned" from the empirical examples instead of "prespecified" by an analyst based on prior knowledge of the datasets (Hagan, et al., 1996).
- The noisy information inevitably included in the examples supplies a trained neural network with the ability to generalize, which makes neural networks robust solutions in the presence of previously unseen, incomplete, or imprecise data. The knowledge in a traditional expert system that must be extracted from knowledgeable experts of a domain area may be subjective and incomplete. This is because the experts may have a biased or even incorrect understanding of reality, they may not be aware of underlying rules they have used, and they may have difficulty articulating these rules. Conversely, knowledge in a neural network is acquired through learning by empirical (real-world) training examples. Although experts are essential in selecting and preparing sample training data, generally their personal biases are excluded from the knowledge acquisition process.
- Knowledge in an expert system is represented by logical rules made up of binary predicates. Numerical attributes have to be converted to binary true/false statements, which may cause a large amount of information to be lost in the simplification process (Gahegan, 2003). On the other hand, a neural network can embrace data in all formats as long as the data are converted to a numeric representation (Brand and Gerritsen, 2004).
- Most rule-based expert systems fail to generalize a predictable inference if an appropriate match with the perfect rules that must be articulated by experts

TABLE 10–3 Characteristics of selected artificial neural networks (based on observations from Ji, 2000; Qiu and Jensen, 2004; Li and Eastman, 2006; Li, 2008; Hu and Weng, 2009; Weng and Hu, 2008; Filippi et al., 2010; Weng, 2012).

Neural Network	Advantages and Disadvantages
Multi-Layer Perceptron (MLP) with Back-Propagation (BP)	• Often consists of one input layer, one hidden layer, and one output layer. • MLP is very sensitive to the structure of the network, especially the number of hidden layers and the number of nodes in the hidden layer. Determining the optimum number of hidden layer nodes is problematic. There is no well-accepted determination method. • MLP training data must contain both presence and absence data and the output must contain both true and false information. • The training stops at a local minimum instead of a global minimum. • The back-propagation (BP) learning process is time-consuming. • BP training process is not consistent and may result in overfitting. • Slow learning compared to SOM. • Classification phase is faster than SOM. • Black-box working method.
Kohonen's Self-Organizing Map (SOM)	• Consists of one input and one output layer. Therefore, the dilemma of determining the size of the hidden layer is avoided. • Not necessary to provide both presence and absence information. Presence information is sufficient. • It is not affected by local minima in the training process. • Learns rapidly. • Classifies slowly. • Classification results are more consistent than MLP. • Different SOM feature map sizes (e.g., 4 × 4, 5 × 5, etc.) have a significant impact on classification accuracy. • The number of samples for each training class needs to be balanced.
Fuzzy ARTMAP	• Requires little intervention during the training phase. • Minimizes overfitting.

cannot be obtained. Conversely, the knowledge in a neural network derived from real-world training examples inevitably contains noise no matter how careful the examples are selected. Therefore, neural networks are good at generalizing both discrete and continuous data and have a capability to interpolate or adapt to the patterns never seen in the training process. Thus, neural networks are tolerant of noise and missing data and attempt to find the best fit for input patterns (Russell and Norvig, 2010).

- Finally, neural networks continuously adjust the weights as more training data are provided in a changing environment. Thus, they continuously *learn*.

Limitations of Artificial Neural Networks

Despite the useful performance of neural networks in image classification, it is usually difficult to explain in a comprehensive fashion the process through which a given decision or output has been obtained from a neural network (Qiu and Jensen, 2004). There exists an inherent inability to represent knowledge acquired by the network in an explicit form with simple "if–then" rules. The rules of image classification and interpretation learned by the neural network are buried in the weights of the neurons of the hidden layers. It is difficult to interpret these weights due to their complex nature. Therefore, a neural network is often accused of being a *black box* (Qiu and Jensen, 2004).

For these reasons, neural networks may not be accepted as trusted solutions in some critical real-world applications. Using neural networks, an analyst might find it difficult to gain an understanding of the problem at hand because of the lack of an explanatory capability to provide insight into the characteristics of the dataset. For the same reason, it is difficult to incorporate human expertise to simplify, accelerate, or improve the performance of image classification; a neural network always has to learn from scratch. For neural networks to be widely applied in complex remote sensing image classification tasks, an explanation capability should eventually be an integral part of the functionality of a trained neural network. A possible solution is to combine neural networks with fuzzy logic so that the knowledge residing in the hidden neurons of the network can be extracted in the form of fuzzy "if–then" rules (Qiu and Jensen, 2004).



References

Aitkenhead, J. J., and I. H. Aalders, 2011, "Automating Land Cover Mapping of Scotland using Expert System and

Knowledge Integration Methods," *Remote Sensing of Environment*, 115:1285–1295.

Akerkar, R., and P. Sajja, 2009, *Knowledge-Based Systems*, London: Jones and Bartlett, 354 p.

Atkinson, P. M., and A. R. L. Tatnall, 1997, "Neural Networks in Remote Sensing," *International Journal of Remote Sensing*, 18(4):699–709.

Brand, E., and R. Gerritsen, 2004, "Neural Networks," *Data Base Mining Solutions*, www.dbmasmag.com/9807m06.html.

Breiman, L., 2001, "Random Forests," *Machine Learning*, 45: 5–32.

Breiman, L., and A. Cutler, 2014, *Random Forests*, http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm.

Brown, M. E., Lary, D. J., Vrieling, A., Stathakis, D., and H. Mussa, 2008, "Neural Networks as a Tool for Constructing Continuous NDVI Time Series from AVHRR and MODIS," *International Journal of Remote Sensing*, 29(24):7141–7158.

Buchanan, B. G., and J. Lederberg, 1971, "The Heuristic DENDRAL Program for Explaining Empirical Data," *IFIP Congress* (1):179–188.

Carpenter, G. A., Gjaja, M. N., Gopal, S., and C. E. Woodcock, 1997, "ART Neural Networks for Remote Sensing: Vegetation Classification from Landsat TM and Terrain Data," *IEEE Transactions on Geoscience and Remote Sensing*, 35(2), 308–325.

Chen, W., Li, S., Wang, Y., Chen, G., and S. Liu, 2014, "Forested Landslide Detection using LiDAR data and the Random Forest Algorithm: A Case Study of the Three Gorges, China," *Remote Sensing of Environment*, 152:291–301.

Clark Labs CTA, 2014, "Classification Tree Analysis (CTA)," Worcester: Clark Labs, 2 p., http://www.clarklabs.org/applications/upload/classification-tree-analysis-idrisi-focus-paper.pdf.

Congalton, R. G., and K. Green, 2009, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd E., Boca Raton, FL: Lewis Publishers, 183 p.

Coops, N. C., Wulder, M. A., and J. C. White, 2006, *Integrating Remotely Sensed and Ancillary Data Sources to Characterize A Mountain Pine Beetle Infestation*, Victoria: The Canadian Forest Service, 33 p.

Darlington, K., 1996, "Basic Expert Systems," *Information Technology in Nursing*, London: British Computer Society, http://www.scism.sbu.ac.uk/~darlink.

DeFries, R. and J. Chan, 2000, "Multiple Criteria for Evaluating Machine Learning Algorithms for Land Cover Classification," *Remote Sensing of Environment*, 74:503–515.

Dreyfus, H. L., and S. E. Dreyfus, 2001, "From Socrates to Expert Systems: The Limits and Dangers of Calculative Rationality," *Selected Papers of Hubert Dreyfus*, Berkeley: Dept. of Philosophy, Regents of the University of California, March, http://ist-socrates.berkeley.edu/~hdreyfus/html/paper_socrates.html.

Duro, D. C., Franklin, S. E., and M. G. Dube, 2012, "A Comparison of Pixel-based and Object-based Image Analysis with Selected Machine Learning Algorithms for the Classification of Agricultural Landscapes using SPOT-5 HRG Imagery," *Remote Sensing of Environment*, 118:259–272.

Exelis ENVI, 2014, *ENVI* Classic *Tutorial: Decision Tree Classification,* Boulder: Exelis Visual Information Solutions, 12 p.

Filippi, A. M., and J. R. Jensen, 2006, "Fuzzy Learning Vector Quantization for Hyperspectral Coastal Vegetation Classification," *Remote Sensing of Environment*, 100:512–530.

Filippi, A. M., Brannstrom, C., Dobreva, I., Cairns, D. M., and D. Kim, 2009, "Unsupervised Fuzzy ARTMAP Classification of Hyperspectral Hyperion Data for Savanna and Agriculture Discrimination in the Brazilian Cerrado," *GIScience & Remote Sensing*, 46(1):1–23.

Filippi, A. M., Dobreva, I., Klein, A. G., and J. R. Jensen, 2010, "Chapter 14: Self-Organizing Map-based Applications in Remote Sensing," in *Self-Organizing Maps*, G. K. Matsopoulos (Ed.,), Vienna: In-Tech, Inc., 231–248.

Foody, G. M., 1996, "Fuzzy Modelling of Vegetation from Remotely Sensed Imagery," *Ecological Modeling*, 85:2–12.

Foody, G. M., and M. K. Arora, 1997, "An Evaluation of Some Factors Affecting the Accuracy of Classification by an Artificial Neural Network," *International Journal of Remote Sensing*, 18(4):799–810.

Foody, G. M., Lucas, R. M., Curran, P. J., and M. Honzak, 1997, "Non-linear Mixture Modelling without End-members Using an Artificial Neural Network," *International Journal of Remote Sensing*, 18:937–953.

Foody, G. M., and A. Mathur, 2004, "A Relative Evaluation of Multiclass Image Classification by Support Vector Machines," *IEEE Transactions on Geoscience and Remote Sensing*, 42:1335–1343.

Foody, G. M., and A. Mathur, 2006, "The Use of Small Training Sets Containing Mixed Pixels for Accurate Hard Image Classification: Training on Mixed Spectral Responses for Classification by SVM," *Remote Sensing of Environment*, 103:179-189.

Foody, G. M., McCulloch, M. B., and W. B. Yates, 1995, "Classification of Remotely Sensed Data by an Artificial Neural Network: Issues Related to Training Data Characteristics," *Photogrammetric Engineering & Remote Sensing*, 61:391–401.

Friedl, M. A., Sulla-Menshe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and X. Huang, 2010, "MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of New Datasets," *Remote Sensing of Environment*, 114:168–182.

Gahegan, M., 2003, "Is Inductive Machine Learning Just Another Wild Goose (or Might It Lay the Golden Egg)?" *International Journal of Geographical Information Science*, 17(1):69–92.

Gonzalez Vilas, L., Spyrakos, E., and J. M. Torres Palenzuela, 2011, "Neural Network Estimation of Chlorophyll *a* from MERIS Full Resolution Data for the Coastal Waters of Galician *rias* (NW Spain)," *Remote Sensing of Environment*, 115:524–535.

Guo, L., Chehata, N., Mallet, C., and S. Boukir, 2011, "Relevance of Airborne Lidar and Multispectral Image Data for Urban Scene Classification using Random Forests, I*SPRS Journal of Photogrammetry & Remote Sensing*, 66:56–66.

Hagan, M. T., Demuth, H. B., and M. Beale, 1996, *Neural Network Design*, Boston: PWS Publishing.

Hayes, M. M., Miller, S. N., and M. A. Murphy, 2014, "High-resolution Landcover Classification using Random Forests," *Remote Sensing Letters*, 5(2):112–121.

Haykin, S., 1994, *Neural Networks: A Comprehensive Foundation*, New York: Macmillan College Publishing, 696 p.

Hexagon ERDAS Imagine, 2014, *Imagine Expert Classifier*, https://wiki.hexagongeospatial.com//index.php?title= Expert _Classification.

Hodgson, M. E., Jensen, J. R., Tullis, J. A. Riordan, K. D., and C. M. Archer, 2003, "Synergistic Use of Lidar and Color Aerial Photography for Mapping Urban Parcel Imperviousness," *Photogrammetric Engineering & Remote Sensing*, 69(9):973–980.

Hu, X., and Q. Weng, 2009, "Estimating Impervious Surfaces from Medium Spatial Resolution Imagery Using the Self-organizing Map and Multi-Layer Perceptron Neural Networks," *Remote Sensing of Environment*, 113:2089–2102.

Huang, C., Davis, L. S., and J. R. Townshend, 2002, "An Assessment of Support Vector Machines for Land Cover Classification," *International Journal of Remote Sensing*, 23(4):725–749.

Huang, X., and J. R. Jensen, 1997, "A Machine-Learning Approach to Automated Knowledge-base Building for Remote Sensing Image Analysis with GIS Data," *Photogrammetric Engineering & Remote Sensing*, 63(10):1185–1194.

Im, J., and J. R. Jensen, 2005, "A Change Detection Model Based on Neighborhood Correlation Image Analysis and Decision Tree Classification," *Remote Sensing of Environment*, 99:326–340.

Im, J., Jensen, J. R., Jensen, R. R., Gladden, J., Waugh, J., and M. Serrato, 2012a, "Vegetation Cover Analysis of Hazardous Waste Sites in Utah and Arizona Using Hyperspectral Remote Sensing," *Remote Sensing*, 2012(4):327–353.

Im, J., Lu, Z., Rhee, J., and L. Quackenbush, 2012b, "Impervious Surface Quantification using a Synthesis of Artificial Immune Networks and Decision/Regression Trees from Multi-sensor Data," *Remote Sensing of Environment*, 117:102–113.

Jackson, P., 1990, *Introduction to Expert Systems*, 2nd Ed., Wokingham, England: Addison-Wesley.

Jensen, J. R., 1978, "Digital Land Cover Mapping Using Layered Classification Logic and Physical Composition Attributes," *American Cartographer*, 5(2):121–132.

Jensen, J. R., 2007, *Remote Sensing of the Environment: An Earth Resource Perspective*, Upper Saddle River, NJ: Prentice-Hall, 590 p.

Jensen, J. R., Im, J., Hardin, P., and R. R. Jensen, 2009, "Chapter 19: Image Classification," in *The Sage Handbook of Remote Sensing*, Warner, T. A., Nellis, M. D. and G. M. Foody, (Eds.), 269–296.

Jensen, J. R., Qiu, F., and M. Ji, 1999, "Predictive Modeling of Coniferous Forest Age Using Statistical and Artificial Neural Network Approaches Applied to Remote Sensing Data," *International Journal of Remote Sensing*, 20(14):2805–2822.

Jensen, J. R., Qiu, F., and K. Patterson, 2001, "A Neural Network Image Interpretation System to Extract Rural and Urban Land Use and Land Cover Information from Remote Sensor Data," *Geocarto International*, 16(1):19–28.

Ji, C. Y., 2000, "Land-Use Classification of Remotely Sensed Data Using Kohonen Self-Organizing Feature Map Neural Networks," *Photogrammetric Engineering & Remote Sensing*, 66(12):1451–1460.

Jiang, D., Yang, X., Clinton, N., and N. Wang, 2004, "An Artificial Neural Network Model for Estimating Crop Yields using Remotely Sensed Information," *International Journal of Remote Sensing*, 25(9):1723–1732.

Jonsson, A., Morris, P., Muscettola, N., Rajan, K., and B. Smith, 2000, "Planning in Interplanetary Space: Theory and Practice," in *Proceedings, 5th International Conference on Artificial Intelligence Planning Systems (AIPS-00)*, Breckenridge, CO: AAAI Press, 177–186.

Khorram, S., Yuan, H., and C. F. Van Der Wiele, 2011, Development of a Modified Neural Network-based Land Cover Classification System using Automated Data Selector and Multiresolution Remotely Sensed Data," *GeoCarto Internationa*l, 26(6):435–457.

Lawrence, R. L., and A. Wright, 2001, "Rule-based Classification Systems Using Classification and Regression Tree (CART) Analysis," *Photogrammetric Engineering & Remote Sensing*, 7(10):1137–1142.

Li, Z., 2008, "Fuzzy ARTMAP Based Neurocomputational Spatial Uncertainty Measures," *Photogrammetric Engineering & Remote Sensing*, 74(12):1573–1584.

Li, Z., and J. R. Eastman, 2006, "Nature and Classification of Unlabelled Neurons in the Use of Kohonen's Self-Organizing Map for Supervised Classification," *Transactions in GIS*, 10(4):599–613.

Liaw, A., and M. Weiner, 2014, *RandomForest: Breiman and Cultler's Random Forests for Classification and Regression*, http://CRAN.R-project.org/package=rpart.

Lu, D., and Q. Weng, 2007, "A Survey of Image Classification Methods and Techniques for Improving Classification Performance," *International Journal of Remote Sensing*, 28, 823–870.

Mannan, B., Roy, J., and A. K. Ray, 1998, "Fuzzy ARTMAP Supervised Classification of Remotely-sensed Images," *International Journal of Remote Sensing*, 19, 767–774.

McIver, D. K., and M. A. Friedl, 2002, "Using Prior Probabilities in Decision-Tree Classification of Remotely Sensed Data," *Remote Sensing of Environment*, 81:253–261.

Merriam-Webster, 2003, *Merriam-Webster Dictionary*, Springfield, MA: Merriam-Webster; http://www.m-w.com/dictionary.htm.

Meyer, D., 2014, *Support Vector Machines*, http://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf.

Meyer, D., 2013, *e1071: Miscellaneous Functions of the Department of Statistics*, http://cran.r-project.org/web/packages/e1071/index.html.

Michalski, R. S., 1983, "A Theory and Methodology of Inductive Learning," in Michalski, R. S., Carbonell, S., and T. M. Mitchell (Eds.), *Machine Learning*, Vol. 1, San Mateo: Morgan Kaufmann Publishers.

Oxford, 2014, *Definition of Artificial Intelligence*, London: Oxford Dictionaries, http://www.oxforddictionaries.com/us/definition/english/artificial-intelligence.

Pal, M., 2005, "Random Forest Classifier for Remote Sensing Classification," *International Journal of Remote Sensing*, 26(1):217.

Pal, M., and P. M. Mather, 2003, "An Assessment of the Effectiveness of Decision Tree Methods for Land Cover Classification," *Remote Sensing of Environment*, 86:554–565.

Pal, M., and P. M. Mather, 2005, "Support Vector Machines for Classification in Remote Sensing," *International Journal of Remote Sensin*g, 26(5):1007–1011.

PC AI, 2002, "Expert Systems," *Personal Computing Artificial Intelligence Electronics Magazine*, Feb. 14, 2002, www.PCAI.com,

Pena-Barragan, J. M., Ngugi, M. K., Plant, R. E., and J. Six, 2011, "Object-based Crop Classification using Multiple Vegetation Indices, Textural Features and Crop Phenology," *Remote Sensing of Environmen*t, 115:1301–1316.

Qiu, F., and J. R. Jensen, 2004, "Opening the Black Box of Neural Networks for Remote Sensing Image Classification," *International Journal of Remote Sensing*, in press.

Quinlan, J. R., 2003, *Data Mining Tools See5 and C5.0*. St. Ives NSW, Australia: RuleQuest Research. http://www.rulequest.com/see5-info.html.

R Development Core Team, 2014, *The R Project for Statistical Computing*, http://www.r-project.org/.

Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., and C. Jeganathan, 2012a, "Random Forest Classification of Mediterranean Land Cover using Multi-seasonal Imagery and Multi-seasonal Texture," *Remote Sensing of Environment*, 121:93–107.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., and J. P. Rigol-Sanchez, 2012b, "An Assessment of the Effectiveness of A Random Forest Classifier for Land-Cover Classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, 67:93–104.

Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C., and D. Roberts, 2008, "Mapping Land-cover Modifications Over Large Areas: A Comparison of Machine Learning Algorithms," *Remote Sensing of Environment*, 112:2272–2283.

RuleQuest Research, Inc., 2014, *C5.0 (Unix/Linux), See5 (for Windows,) and Cubist*, Australia: RuleQuest Research Pty Ltd., http://www.rulequest.com/.

Russell, S. J., and P. Norvig, 2010, *Artificial Intelligence: A Modern Approach*, 3rd Ed., Upper Saddle River, NJ: Prentice-Hall, 1109 p.

Salah, M., Trinder, J., and A. Shaker, 2009, "Evaluation of the Self-Organizing Map Classifier for Building Detection from Lidar Data and Multispectral Aerial Images," *Spatial Science*, 54(2):1–20.

Salford Systems, 2014a, *CART Classification and Regression Trees*, San Diego: Salford Systems, Inc., https://www.salford-systems.com/products/cart.

Salford Systems, 2014b, *Random Forests*, San Diego: Salford Systems, Inc., https://www.salford-systems.com/products/randomforests#random-forests.

Santi, E., Pettinato, S., Paloscia, S., Pampaloni, P., Fontanelli, G., Crepaz, A., and M. Valt, 2014, "Monitoring Alpine Snow using Satellite Radiometers and Artificial Neural Networks," *Remote Sensing of Environment*, 144:179–186.

Schneider, A., 2012, "Monitoring Land Cover Change in Urban and Peri-urban Areas using Dense Time Stacks of Landsat Satellite Data and a Data Mining Approach," *Remote Sensing of Environment*, 124:689–704.

Sexton, J. O., Urban, D. L., Donohue, M. J., and C. Song, 2013, "Long-term Land Cover Dynamics by Multi-temporal Calcification Across the Landsat-5 Record," *Remote Sensing of Environment*, 128:246–258.

Stow, D., Coulter, L., Kaiser, J., Hope, A., Service, D., Schutte, K., and A. Walters, 2003, "Irrigated Vegetating Assessments for Urban Environments," *Photogrammetric Engineering and Remote Sensing*, 69(4):381–390.

Su, L., Chopping, M. J, Rango, A., Martonchik, H. V., and D. P. Peter, 2006, "Support Vector Machines for Recognition of Semi-arid Vegetation Types using MISR Multi-angle Imagery," *Remote Sensing of Environment*, 107: 299–311.

Swain, P. H., and H. Hauska, 1977, "The Decision Tree Classifier: Design and Potential," *IEEE Transactions on Geoscience and Remote Sensing*, 15:142–147.

Therneau, T., Atkinson, B., and B. Ripley, 2014, *rpart: Recursive Partitioning and Regression Trees*, http://CRAN.R-project.org/package=rpart.

TIBCO S-Plus, 2014, *Spotfire S-PLUS*®, Sommerville, MD: TIBCO, http://www.solutionmetrics.com.au/products/splus/S-PLUSBrochure.pdf

Tso, B., and P. M. Mather, 2001, *Classification Methods for Remotely Sensed Data*, New York: Taylor & Francis, 332 p.

Tullis, J. A., and J. R. Jensen, 2003, "Expert System House Detection in High Spatial Resolution Imagery Using Size, Shape, and Context," *Geocarto International*, 18(1):5–15.

Turing, A. M., 1950, "Computing Machinery and Intelligence," *Mind*, 59:439–460.

Van der Linden, S., and P. Hostert, 2009, "The Influence of Urban Structures on Impervious Surface Maps from Airborne Hyperspectral Data," *Remote Sensing of Environment*, 113:2298–2305.

Wang, J. G., Neskovic, P., and L. N. Cooper, 2005, "Training Data Selection for Support Vector Machines," *Lecture Notes in Computer Science*, 3600:554–564.

Weng, Q., 2012, "Remote Sensing of Impervious Surfaces in the Urban Areas: Requirements, Methods, and Trends," *Remote Sensing of Environmen*t, 117:34–49.

Weng, Q., and X. Hu, 2008, "Medium Spatial Resolution Satellite Imagery for Estimating and Mapping Urban Impervious Surfaces Using LSMA and ANN," *IIEE Transactions on Geoscience and Remote Sensing*, 46(8):2397–2406.

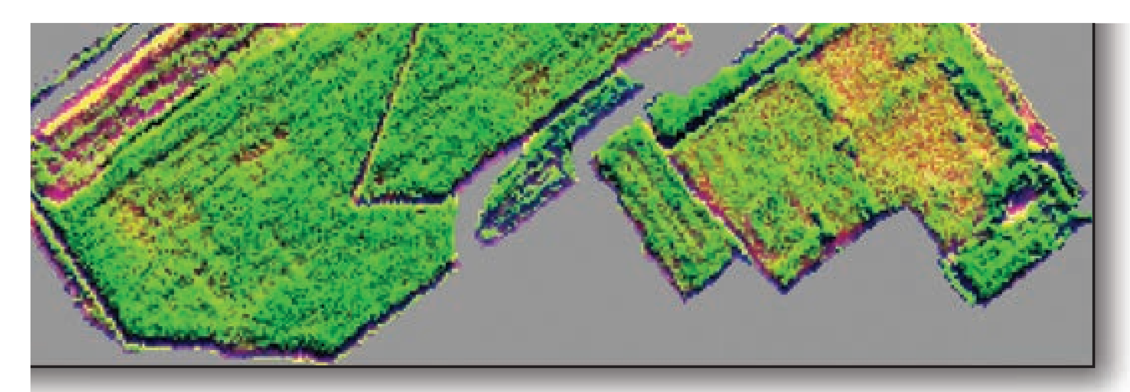
Xian, G., Homer, C., and J. Fry, 2009, "Updating the 2001 National Land Cover Database Land Cover Classification using Landsat Imagery Change Detection Methods," *Remote Sensing of Environment*, 113:1133–1147.

Xian, G., Homer, C., Dewitz, J., Fry, J., Hossain, N., and J. Wickham, 2011, "Change of Impervious Surface Area Between 2001 and 2006 in the Conterminous United States," *Photogrammetric Engineering & Remote Sensing*, 77(8):758–762

Yang, C., and P. Chung, 2002, "Knowledge-based Automatic Change Detection Positioning System for Complex Heterogeneous Environments," *Journal of Intelligent and Robotic Systems*, 33:85–98.

Zhang, Q., and J. Wang, 2003, "A Rule-based Urban Land Use Inferring Method for Fine-resolution Multispectral Imagery," *Canadian Journal of Remote Sensing*, 29(1):1–13.

11 INFORMATION EXTRACTION USING IMAGING SPECTROSCOPY

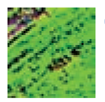


Source of AVIRIS data: NASA

Imaging spectrometry is:

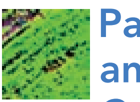
> the simultaneous acquisition of images in many relatively narrow, contiguous and/or noncontiguous spectral bands throughout the ultraviolet, visible, and infrared portions of the electromagnetic spectrum.

The dense set of spectral measurements is collected using an imaging spectrometer, commonly referred to as a **hyperspectral** remote sensing instrument. Many, although not all, surface materials such as vegetation, water, rocks, etc., have diagnostic absorption features that are only 10 – 20 nm wide. Therefore, hyperspectral sensors that acquire data in many contiguous or noncontiguous 10 – 20-nm wide bands throughout the 400 – 2,500-nm region of the electromagnetic spectrum sometimes capture spectral data with sufficient resolution for the direct identification of those materials (Im and Jensen, 2008). The value of using a hyperspectral imaging spectrometer lies in its ability to provide a high-resolution reflectance (or emittance) spectrum for each pixel in the image (Goetz et al., 1985; Schaepman et al., 2009; Chiu et al., 2011).



Overview

This chapter first reviews the differences between panchromatic, multispectral, and hyperspectral remote sensing instruments. Because of the unique nature of the hyperspectral data, special-purpose digital image processing algorithms are used to extract useful information. The pre-processing steps required to analyze hyperspectral data are reviewed, including image quality assessment, radiometric calibration to scaled surface reflectance, geometric correction and dimensionality reduction. Special-purpose information extraction techniques are reviewed, including: the identification of spectrally-pure endmembers, spectral angle mapping, subpixel classification (linear spectral unmixing), spectral library matching techniques, the use of various hyperspectral indices, and derivative spectroscopy.



Panchromatic, Multispectral and Hyperspectral Data Collection

Optical remote sensing systems typically collect panchromatic, multispectral, and hyperspectral data within the region from 400 – 2,500 nm. There are also now hyperspectral thermal-infrared sensors.

Panchromatic

Some remote sensing systems record reflected or emitted electromagnetic energy in a single broad wavelength interval, e.g., from 500 to 700 nm (includes green and red wavelength energy) or 500 to 900 nm (includes green, red, and near-infrared energy). This is typically called a *panchromatic* band or panchromatic remote sensing. Considerable panchromatic (black-and-white) aerial photography has been collected through the years and used extensively in stereoscopic photogrammetric applications. Several satellite remote sensing systems have only one panchromatic band (e.g., WorldView-1 from 400 – 900 nm at 0.5 × 0.5 m spatial resolution; EROS B from 500 – 900 nm at 0.7 × 0.7 m). Numerous multispectral sensing systems have a

single panchromatic band, which is often at much higher spatial resolution than the other data collected (e.g., Landsat ETM$^+$, Landsat 8, SPOT 5, GeoEye-1).

Multispectral

Most satellite and airborne *multispectral* remote sensing systems collect data in 3 to 10 spectral bands with relatively broad bandwidths (e.g., natural color and color-infrared digital frame cameras, Landsat TM, ETM$^+$, Landsat 8, SPOT 6, GeoEye-1, WorldView-2, QuickBird, Pleiades).

Hyperspectral

Hyperspectral remote sensing systems typically collect data in >10 spectral bands with relatively narrow bandwidths (e.g., Gao et al., 2009; Thenkabail et al., 2011). *However, there is no agreed upon minimum number of bands or bandwidth dimension required for a dataset to be considered hyperspectral.* The collection of spectral reflectance and emittance information in hundreds of spectral bands requires different approaches to sensor system design. Two basic approaches to imaging spectrometry are shown in Figure 11-1. The "whiskbroom" scanner linear array approach is analogous to the scanner approach used for Landsat ETM$^+$, except that radiant flux from within the IFOV is passed onto a spectrometer, where it is dispersed and focused onto a linear array consisting of >10 detectors (Figure 11-1a). The terrain within each IFOV (i.e., pixel) is sensed in as many spectral bands as there are detector elements in the linear array.

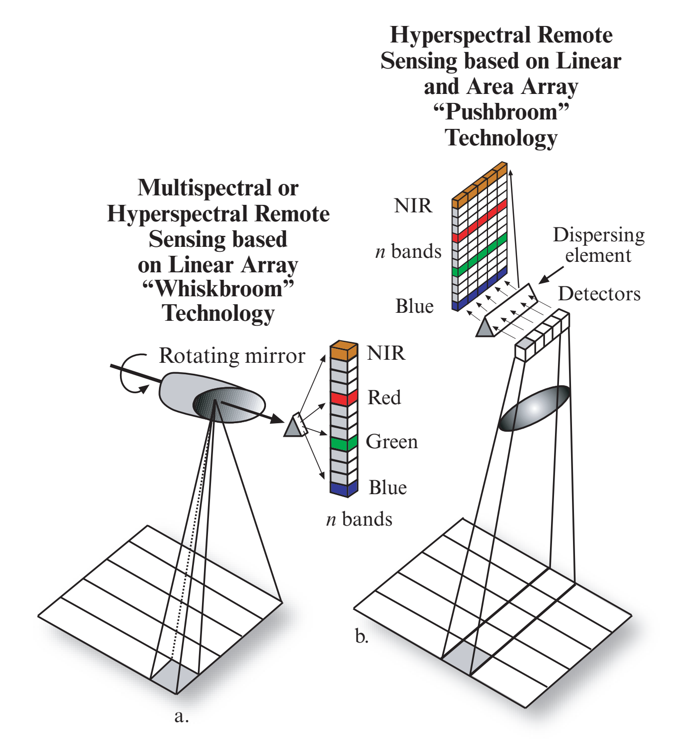


FIGURE 11–1 a) A multispectral or hyperspectral "whiskbroom" remote sensing system based on linear array detector technology and a scanning mirror (e.g., AVIRIS). b) A hyperspectral instrument based on linear and area array "pushbroom" technology.

Other imaging spectrometers make use of linear and two-dimensional area arrays of detectors. In this situation, there is a dedicated column of spectral detector elements for each cross-track pixel in the scene (Figure 11-1b). This "pushbroom" configuration improves image geometry and radiometry because a scanning mirror is not used, allowing each detector to dwell longer on an IFOV, thereby resulting in a more accurate recording of the radiant flux as it exits the landscape.

Satellite Hyperspectral Sensors

Satellite hyperspectral remote sensing systems include (Table 11-1):

- NASA's Moderate Resolution Imaging Spectrometer (MODIS) onboard the *Terra* and *Aqua* satellites, collects data in 36 co-registered spectral bands: 20 bands from 0.4 – 3 μm and 16 bands from 3 – 15 μm.
- NASA's EO-1 Hyperion sensor with 198 useful bands out of 220 that are calibrated, including channels 8 – 57 for the VNIR and channels 77 – 224 for the SWIR extending from 400 – 2,500 nm with a spectral resolution of 10 nm.
- European Space Agency (ESA) Medium Resolution Imaging Spectrometer (MERIS) on Envisat used "pushbroom" technology to collect 15 selectable bands in the 390 – 1,040 nm region with bandwidths programmable between 2.5 and 30 nm. MERIS is no longer operational (ESA MERIS, 2014).
- NASA's Hyperspectral Infrared Imager (HyspIRI) is scheduled to include two instruments (NASA HyspIRI, 2014): 1) an imaging spectrometer that senses in the visible to short-wave infrared (VSWIR: 380 – 2,500 nm) in 10 nm contiguous bands, and 2) a multispectral imager that collects data in 8 bands in the thermal-infrared (TIR) region; one band in the region from 3 – 5 μm to measure hot targets and seven bands in the region from 7 – 12 μm (Ramsey et al., 2012). The VSWIR and TIR sensors will have a spatial resolution of 60 × 60 m.

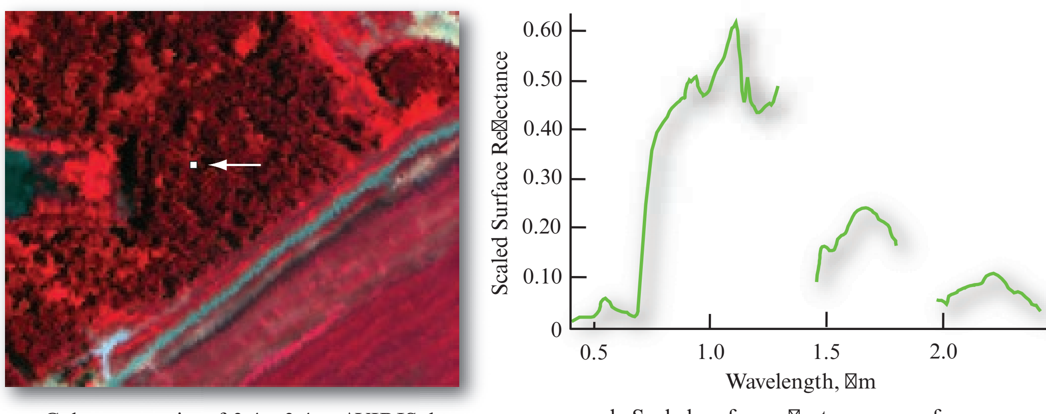
Additional information about these satellite sensor systems is found in Chapter 2.

Airborne Optical Hyperspectral Sensors

Selected *suborbital airborne* hyperspectral remote sensing systems include (Table 11-1):

- NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) uses a whiskbroom scanning

Imaging Spectrometry-Derived Spectra



a. Color composite of 3.4 × 3.4 m AVIRIS data (RGB = bands 50, 30, 20) from Run 08 Scene 05.

b. Scaled surface reflectance curve for a pixel of loblolly pine (*Pinus taeda*).

FIGURE 11–2 a) A color composite image of Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) imagery obtained on July 26, 1999, at the Savannah River Site near Aiken, SC, at 3.4 × 3.4 m spatial resolution. Original imagery courtesy of NASA. b) Scaled surface reflectance spectrum for a pixel of loblolly pine (*Pinus taeda*) forest at the location shown in (a).

mirror and linear arrays of silicon (Si) and indium-antimonide (InSb) to record energy in 224 bands, each 10 nm wide in the 400 – 2,500 nm region

- Compact Airborne Imaging Spectrometer 1500 (CASI-1500) is a pushbroom imaging spectrometer that collects data in the region from 380 – 1,050 nm and has a 40° total field of view across 1,500 pixels. Unlike most other hyperspectral sensors, CASI-1500 is spectrally programmable meaning that the user can specify which bands are to be collected for a specific application (ITRES CASI-1500, 2014).
- SASI-600 collects 100 bands of SWIR hyperspectral data from 950 – 2,450 nm at 15-nm intervals with 600 across-track pixels. SASI-600 data are especially useful for geological exploration and plant speciation (ITRES SASI-600, 2014).
- HyVista HyMap is a whiskbroom hyperspectral sensor with 128 bands from 450 – 2,480 nm with contiguous spectral coverage (except in the atmospheric water vapor bands near 1,400 and 1,900 nm) with average spectral sampling intervals between 13 – 17 nm (HyVista Hymap, 2014).
- DAIS 7915 (DLR), manufactured by GER, Inc., was a 79-band whiskbroom sensor sensitive to the VIS-NIR-SWIR-TIR regions from 400 – 1,300 nm (Ben-Dor et al., 2002; GAC, 2013).
- AISA Eagle hyperspectral sensor records 244–488 bands in the region from 380 – 2,500 nm with a 3.3 nm average spectral sampling interval (Specim AISA, 2014).
- AISA FENIX hyperspectral sensor records 620 bands in the region from 380 – 2,500 nm with a 6 – 7 nm average spectral sampling interval (Specim AISA, 2014).

Airborne Thermal-Infrared Hyperspectral Sensors

ITRES Research, Ltd., markets mid-wavelength infrared (MASI-600) and thermal-infrared (TASI-600) hyperspectral instruments (Table 11-1):

- MASI-600 is an airborne mid-wavelength thermal-infrared hyperspectral sensor that records 64 bands in the 4 – 5 μm range with a 40° across-track field of view encompassing 600 pixels (ITRES MASI-600, 2014).
- TASI-600 is a pushbroom hyperspectral thermal-infrared sensor system that collects 32 bands in the region from 8 to 11.5 μm with a 40° across-track field of view encompassing 600 pixels (ITRES TASI-600, 2014).

Traditional broadband remote sensing systems such as Landsat TM and SPOT HRV *undersample* the information available from a reflectance spectrum by making only a few measurements in spectral bands sometimes several hundred nanometers wide. Conversely, imaging spectrometers sample at close intervals (bands typically on the order of 10 – 20 nm wide) and have a sufficient number of spectral bands to allow construction of *spectra* that closely resemble those acquired by laboratory spectroradiometers. Analysis of imaging spectrometer data allows the extraction of a detailed spectrum for each picture element in the image. Such spectra may allow direct identification of specific materials within the IFOV of the sensor based upon their reflectance/absorptance characteristics, including asphalt, concrete, rocks and minerals, atmospheric gases, vegetation, snow and ice, and dissolved matter in water bodies (e.g., Nolin and Dozier, 2000; Chiu et al., 2011). For example, Figure 11-2 depicts the

TABLE 11–1 Characteristics of selected hyperspectral remote sensing systems (Dalponte et al., 2009; DLR ROSIS, 2014; HyVista HyMap, 2014; ITRES, 2014; Plaza et al., 2009; Specim AISA, 2014; NASA, and others).

Sensor	Source	Platform	Number of Bands	Maximum Spectral Resolution (nm)	Spectral Range (μm)
EO-1 Hyperion	NASA Goddard SFC	Satellite	220	10	0.4–2.5
MODIS	NASA Goddard SFC	Satellite	36	40	0.4–14.3
MERIS	European Space Agency	Satellite	15	2.5–30	0.390–1.04
CRIS Proba	European Space Agency	Satellite	up to 63	1.25	0.415–1.05
AVIRIS	NASA Jet Propulsion Lab	Airborne	224	10	0.4–2.5
HYDICE	Naval Research lab	Airborne	210	7.6	0.4–2.5
PROBE-1	Earth Search Sciences	Airborne	128	12	0.4–2.45
CASI-1500	ITRES Research Ltd.	Airborne	288	2.5	0.4–1.05
SASI-600	ITRES Research Ltd.	Airborne	100	15	0.95–2.45
TASI-600	ITRES Research Ltd.	Airborne	32	250	8–11.5
HyMap	HyVista, Inc.	Airborne	128	15	0.44–2.5
ROSIS	DLR	Airborne	115	4	0.4–0.9
EPS-H	GER, Inc.	Airborne	133	0.67	0.43–12.5
EPS-A	GER, Inc.	Airborne	31	2.3	0.43–12.5
DAIS 7915	GER, Inc.	Airborne	79	15	0.43–12.3
AISA Eagle	Specim, Inc.	Airborne	244–488	3.3	0.4–0.97
AISA Owl	Specim, Inc.	Airborne	96	100	7.7–12.3
AISA Hawk	Specim, Inc.	Airborne	320	8.5	0.99–2.45
AISA Dual	Specim, Inc.	Airborne	500	2.9	0.4–2.45
AISA FENIX	Specim, Inc.	Airborne	620	3.5 and 12	0.38–2.5
MIVIS	Daedalus	Airborne	102	20	0.43–12.7
AVNIR	OKSI	Airborne	60	10	0.43–1.03

scaled surface reflectance characteristics of hyperspectral data obtained over a loblolly pine (*Pinus taeda*) forest on the Savannah River Site near Aiken, SC, in the wavelength interval from 400 – 2,500 nm.



Steps to Extract Information from Hyperspectral Data

Scientists have developed digital image processing methods that may be applied in a relatively straightforward sequence to extract useful information from hyperspectral remote sensor data (e.g., Eismann et al., 2012; Chang et al., 2013). Many of the operations identified in Figure 11-3 are discussed in this chapter.

The analysis of hyperspectral data usually requires sophisticated digital image processing software (e.g., ENVI, ERDAS Imagine; IDRISI, PCI Geomatica, VIPER Tools). This software must be able to:

1. Calibrate (convert) the raw hyperspectral at-sensor radiance to apparent surface reflectance and ideally to surface reflectance. This requires the removal of atmospheric attenuation, topographic effects due to slope and aspect, and any sensor system–induced electronic anomalies.

General Steps Used to Extract Information from Hyperspectral Data

State the nature of the information extraction problem.
* Specify the geographic region of interest.
* Define the classes or biophysical materials of interest.

Acquire appropriate remote sensing and initial ground reference data.
* Select remotely sensed data based on the following criteria:
  - Remote sensing system considerations:
    - Spatial, spectral, temporal, and radiometric resolution
  - Environmental considerations:
    - Atmospheric, soil moisture, phenological cycle, etc.
* Obtain initial ground reference data based on:
  - *a priori* knowledge of the study area

Process hyperspectral data to extract thematic information.
* Subset the study area from the hyperspectral data flight line(s).
* Conduct initial image quality assessment:
  - Visual individual band examination
  - Visual examination of color composite images
  - Animation
  - Statistical individual band examination; signal-to-noise ratio
* Radiometric correction:
  - Collect necessary *in situ* spectroradiometer data (if possible)
  - Collect *in situ* or environmental data (e.g., using radiosondes)
  - Perform pixel-by-pixel correction (e.g., ACORN, FLAASH)
  - Perform band-by-band spectral polishing
  - Empirical line calibration
* Geometric correction/rectification:
  - Use onboard navigation and engineering data (GPS and Inertial Navigation System information)
  - Nearest-neighbor resampling
* Reduce the dimensionality of hyperspectral dataset:
  - Minimum Noise Fraction (MNF) transformation
* End-member determination—locate pixels with relatively pure spectral characteristics:
  - Pixel Purity Index (PPI) mapping
  - N-dimensional end-member visualization
* Methods of mapping and matching using hyperspectral data:
  - Spectral Angle Mapper (SAM)
  - Subpixel classification (spectral mixture analysis)
  - Spectroscopic library matching techniques
  - Matched filter or mixture-tuned matched filter
  - Machine-learning (e.g.,Support Vector Machine, Regression Tree)
  - Indices developed for use with hyperspectral data
  - Derivative spectroscopy

Perform accuracy assessment (Chapter 13).
* Select method:
  - Qualitative confidence-building
  - Statistical measurement
* Determine number of observations required by class.
* Select sampling scheme.
* Obtain ground reference test information.
* Create and analyze error matrix:
  - Univariate and multivariate statistical analysis

Accept or reject previously stated hypothesis.
Distribute results if accuracy is acceptable.

FIGURE 11–3 A generalized flow chart useful for the extraction of quantitative biophysical and thematic information from hyperspectral remote sensor data.

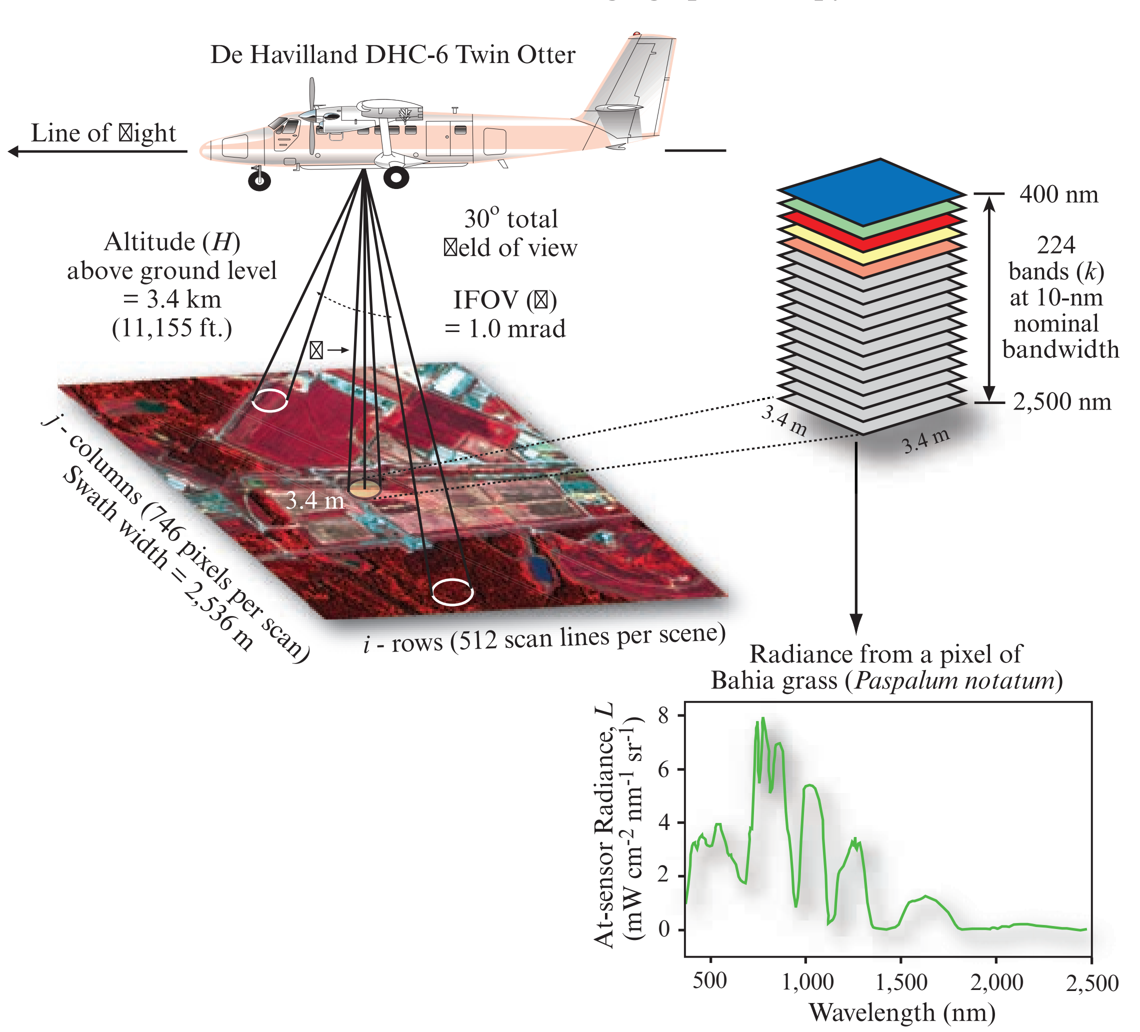


FIGURE 11–4 Conceptual diagram of low-altitude imaging spectroscopy performed using the NASA Jet Propulsion Laboratory's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). The scanner mirror focuses radiant flux onto linear arrays that contain 224 detector elements with a spectral sensitivity ranging from 400–2,500 nm. A spectrum of radiance (*L*) is obtained for each picture element. The radiance data can be processed to yield percent reflectance information. Most AVIRIS overflights obtain data at 20 × 20 m pixels. Sometimes the AVIRIS sensor is flown on other aircraft and at different altitudes above ground level. This particular AVIRIS scene of the Savannah River Site near Aiken, SC, was acquired at 11,155 ft. above sea level in a De Havilland DHC-6 Twin Otter on July 26, 1999. This resulted in 3.4 × 3.4 m spatial resolution data. The at-sensor radiance (*L*) for a pixel of Bahia grass is shown. The image displayed is a color composite of bands 50 (near-infrared), 30 (red), and 20 (green). AVIRIS original imagery courtesy of NASA.

2. Analyze the calibrated remote sensing–derived scaled surface reflectance data to determine its constituent materials by a) deriving endmembers from the hyperspectral data itself, b) comparing *in situ* spectroradiometric data obtained in the field at the time of the overflight with hyperspectral image characteristics, and/or c) comparing hyperspectral image characteristics with a library of laboratory-based spectra such as that provided by Johns Hopkins University, the U.S. Geological Survey, and NASA's Jet Propulsion Laboratory (Baldridge et al., 2009; ASTER Spectral Library, 2014).

To understand how to analyze hyperspectral data, it is useful to examine the image processing methods in the context of real-world applications. The chapter first describes how AVIRIS remote sensor data were analyzed to detect possible stress associated with Bahia grass–covered clay caps on the U.S. Department of Energy Savannah River Site (SRS) near Aiken, SC. AVIRIS sensor system characteristics are summarized in Table 11-1 and Figure 11-4. In 1999, the AVIRIS sensor was also flown in low-altitude mode on a NOAA De Havilland DHC-6 Twin Otter aircraft that acquired data at an altitude of 11,155 ft AGL. This resulted in 224

bands of hyperspectral data with a nominal spatial resolution of approximately 3.4 × 3.4 m. Figure 11-4 depicts the at-sensor radiance (*L*) associated with a single pixel of Bahia grass.

High spatial resolution AVIRIS data of grassland in Monticello, UT, is then analyzed to classify land cover and predict continuous variables (e.g., LAI). Finally, AISA Eagle hyperspectral data of forest treatment plots on the SRS are analyzed to measure biophysical variables (LAI, biomass) and vegetation chemical characteristics (e.g., N, Ca).



Select Study Area from Flight Lines

Both public (e.g., NASA Jet Propulsion Laboratory) and commercial (e.g., HyMap, CASI-1500) data providers usually pre-process the hyperspectral data before transferring it to the user for image analysis (Prost, 2013). This preliminary preprocessing is performed so that the user can look at (browse) the content of the imagery collected and then select a subset for detailed analysis. For example, personnel at NASA's Jet Propulsion Laboratory Data Facility preprocess the raw digital AVIRIS data to remove basic radiometric and geometric errors. The preliminary geometric corrections are applied to each flight line (often referred to as a *run*) of AVIRIS data. Each AVIRIS flight line is then subset into individual scenes. The individual scenes may be viewed via the Internet as quick-look image files. For example, scenes 04, 05, and 06 extracted from run 08 collected over the Savannah River Site on July 26, 1999, are shown in Figure 11-5. There is no overlap of the successive scenes. Therefore, it is relatively straightforward to mosaic the scenes into a single flight line. A datacube of run 08 scene 05 is shown in Figure 11-5d. The dark regions of the 224-band datacube correspond to atmospheric absorption bands present especially at 1,400 and 1,900 nm.



Initial Image Quality Assessment

It is important to perform an initial assessment of the quality of the hyperspectral data. A number of methods can be used to assess image quality including visual examination of multiple-band color composites, inspection of individual bands of imagery, viewing an animated movie of the individual bands, and statistical evaluation of individual bands including a quantitative evaluation of individual band signal-to-noise ratios (SNRs).

Visual Examination of Hyperspectral Color Composite Images

One of the most useful ways to determine if hyperspectral imagery has information of value is to select representative bands from the near-infrared (e.g., AVIRIS band 50), red (e.g., band 30), and green (e.g., band 20) portions of the spectrum and create a false-color composite of the scene as shown in Figure 11-5. Hopefully, the individual bands are co-registered and contain spectral information of value. The analyst can select any color composite of three bands to evaluate. Looking at numerous three-band color composites from the 224 possible bands can be a tedious process. However, the exercise does provide valuable qualitative information about the individual scenes and bands in the hyperspectral dataset. Sometimes this is the only way to determine if a band is geometrically misregistered (offset) when compared to the other bands.

Visual Individual Band Examination

Although it is a tedious process, sometimes there is simply no substitute for painstakingly viewing each band of data to judge its quality. Many bands appear crisp and sharp with good grayscale tonal variation. For example, Figure 11-6a depicts AVIRIS band 50 (centered at 0.8164 μm) from run 08 scene 05. Compare this image with images of AVIRIS band 112 (1.405 μm) and band 163 (1.90 μm) shown in Figure 11-6bc. These images were obtained in two of the prominent atmospheric absorption windows. Therefore, it is not surprising that they are of poor quality. The analyst normally keeps a list of bands that exhibit considerable atmospheric noise (i.e., a bad bands list). Such bands are often deleted from further analysis, as described in subsequent sections. However, it is important to remember that the information in the atmospheric absorption bands may be used by certain atmospheric correction algorithms.

It is very important to determine early in a project if certain bands or particular regions within a given band contain null data values (e.g., values of –9,999) or if there are serious line dropouts (i.e., an entire line has a value of –9,999). Also, an examination of the individual bands and color composites provides information on whether there are serious geometric problems with the data such as extreme image warping. Deep notches along the edges of a scene often indicate spatial image distortion, which must be corrected during the geometric rectification phase of the project.

Animation

Hyperspectral datasets usually contain hundreds of bands. Therefore, most digital image processing systems designed for hyperspectral analysis have an image

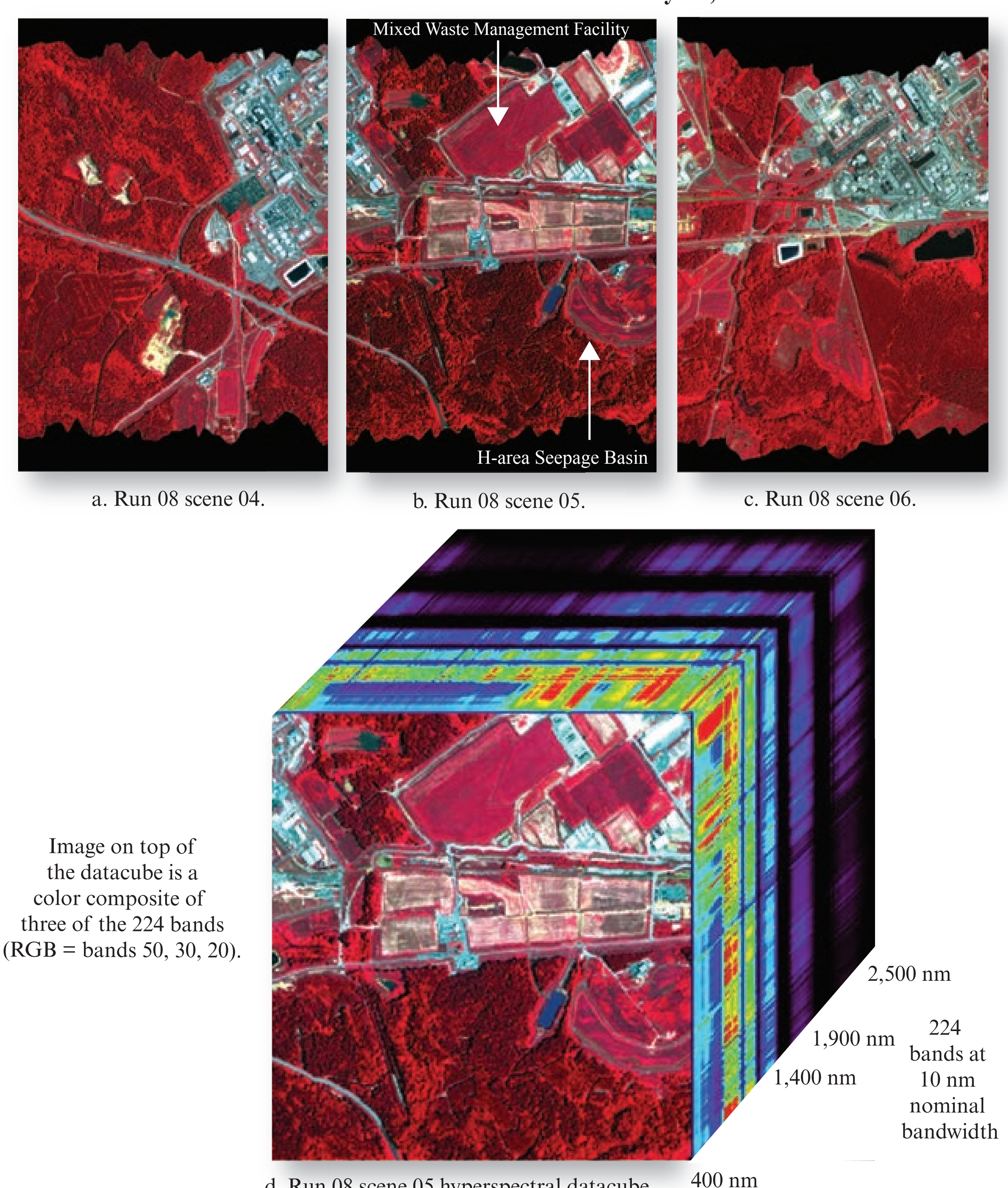


FIGURE 11–5 a–c) Color-infrared color-composites of three AVIRIS scenes (04, 05, and 06) from run 08. The color composites represent a display of just three of the 224 AVIRIS bands (RGB = bands 50, 30, 20). The AVIRIS sensor was flown onboard a De Havilland DHC-6 Twin Otter aircraft at 11,150 ft. (3.4 km) above sea level. This resulted in pixels that were 3.4 × 3.4 m. Each scene is 512 × 746 pixels in dimension. The data were preprocessed at JPL to remove fundamental geometric and radiometric errors. Pixels in the black border regions contain null values (–9,999). AVIRIS imagery courtesy of NASA. d) Hyperspectral datacube of run 08 scene 05. All 224 bands are displayed. Dark areas in the datacube represent atmospheric absorption bands.

Assessing the Quality of Individual AVIRIS Scenes Band by Band

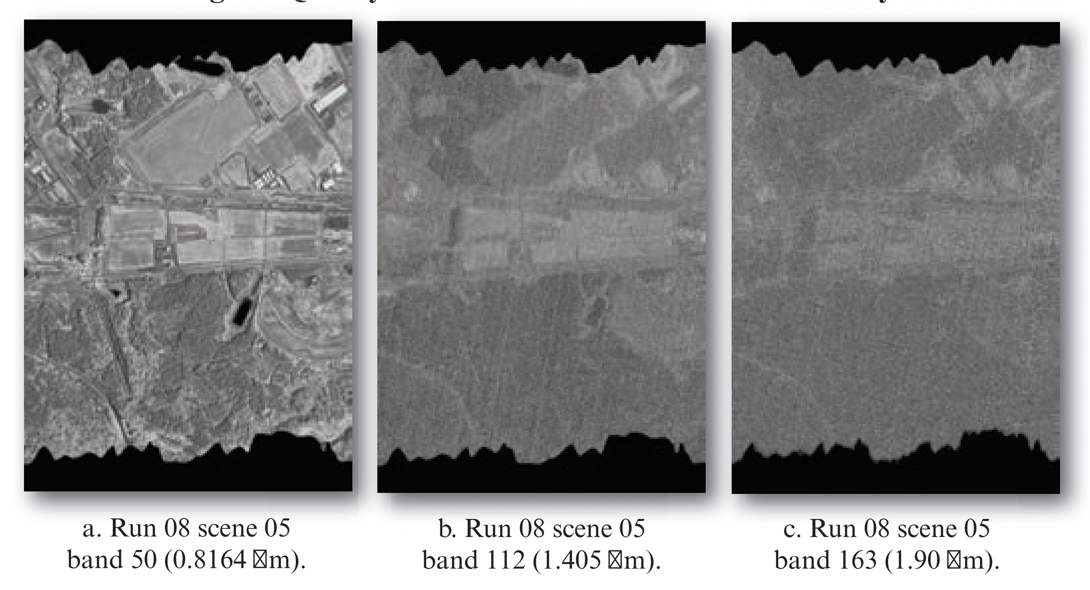


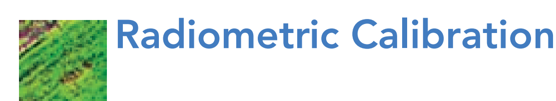
FIGURE 11–6 Selected individual bands of AVIRIS data from run 08 scene 05. a) An image of band 50 after contrast stretching. b) A contrast-stretched image of band 112. c) A contrast-stretched image of band 163. The near-infrared band 50 image is relatively free of attenuation, while bands 112 and 163 exhibit the deleterious effects of atmospheric attenuation.

animation function whereby the analyst selects a certain time rate at which individual bands are displayed on the screen, e.g., every 5 seconds. Examination of hyperspectral bands in this manner allows the analyst to 1) identify individual bands that have serious atmospheric attenuation or electronic noise problems, as previously discussed, and 2) determine if any misregistration (offset) of bands exists.

Occasionally, linear features in the scene such as roads or the edges of water bodies may appear to shift slightly when you are viewing the animated bands. This does not necessarily indicate that the bands are misregistered, but rather that the detectors in one band are more or less sensitive than those in an adjacent band. Sometimes with AVIRIS data there is a slew effect in the detector array readout, which translates to a shift of the IFOV in areas of high spatial frequency. If the displacement appears to be relatively severe, then a more detailed examination should be conducted. This might entail creating color composites of multiple bands to determine if there is actually band-to-band misregistration.

Statistical Individual Band Examination

Examination of the univariate statistics (e.g., mean, median, mode, standard deviation, range) of the individual bands can be helpful when assessing image quality. For example, if a band's 16-bit brightness values are constrained to a very small range with a very small standard deviation, this may indicate a serious problem. When a band is believed to have problems, there is no substitute for analyzing its univariate statistics and histogram. In order to detect many absorption features associated with materials found in nature, a noise level that is approximately an order of magnitude smaller than the absorption depth is required. Therefore, evaluation of each band's signal-to-noise ratio is of value.



Radiometric Calibration

As discussed in Chapter 9, many remote sensing investigations do not concern themselves with radiometric correction of the data to remove the deleterious effects of atmospheric absorption and scattering. However, to use hyperspectral remote sensor data properly, it is generally accepted that the data must be radiometrically corrected. In addition to removing atmospheric effects, this process normally involves transforming the hyperspectral data from at-sensor radiance, L_S (μW cm^{-2} nm^{-1} sr^{-1}) to scaled surface reflectance (Mustard et al., 2001). This allows the remote sensor–derived spectral reflectance data (often referred to as *spectra*) to be quantitatively compared with *in situ* spectral reflectance data obtained on the ground using a handheld spectroradiometer (Figure 11-7a) or with laboratory-derived spectra. Such spectroradiometric data are often stored in spectral libraries (e.g., Baldridge, et al., 2009).

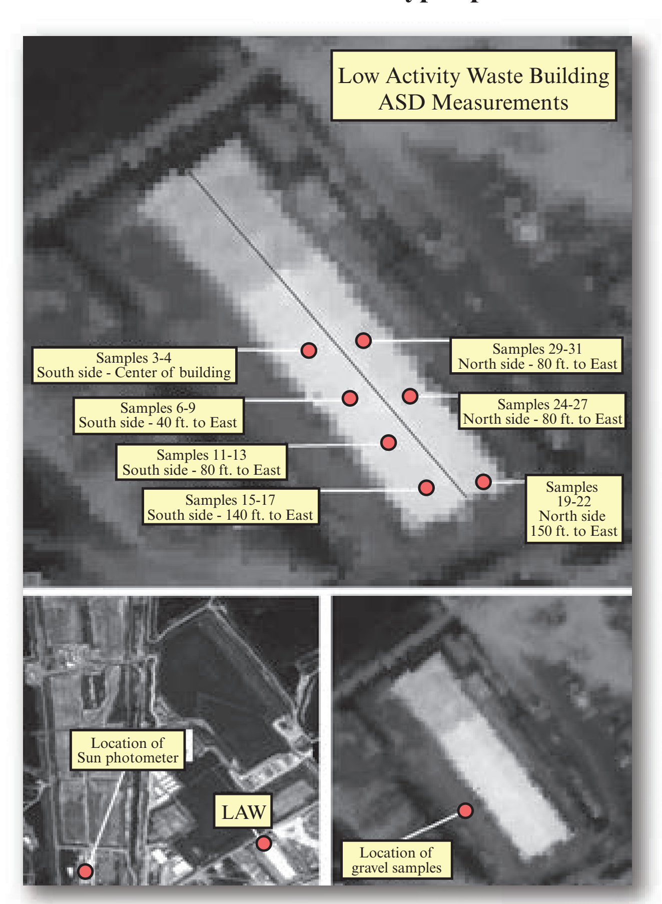
***In Situ* Spectroradiometer Measurements used to Calibrate Hyperspectral Data**



a. Spectroradiometer being calibrated.



b. Calibration panels.



c. Location of spectroradiometer measurements on a rooftop, a gravel site next to the building, and the location of a nearby Sun photometer.

FIGURE 11–7 *In situ* spectroradiometer measurements can be valuable when used to perform absolute atmospheric correction. They are especially useful for empirical line calibration and as *boost spectra*. a) A spectroradiometer being calibrated using a Spectralon plate prior to a Bahia grass measurement. b) Calibration panels placed in the study area. c) The location of additional *in situ* spectroradiometer measurements obtained on top of the Low Activity Waste (LAW) building on the Savannah River Site and adjacent to the building (photographs courtesy of Jensen and Hadley, 2003).

In Situ Data Collection

Whenever possible, it is desirable to obtain handheld *in situ* spectroradiometer measurements on the ground features of interest at approximately the same time as the remote sensing overflight (McCoy, 2005). These measurements should be made with well-respected spectroradiometers (e.g., Spectron Engineering, Analytical Spectral Devices) that cover the same spectral range as the hyperspectral remote sensing system (e.g., 400 – 2,500 nm). Ideally, several calibrated spectroradiometers are used to collect data of the most important materials in the study area at approximately the same time of day and under the same atmospheric conditions as the remote sensing data collection. If this is not possible, then data should be collected at the same time of day on days before and/or after the remote sensing data collection. Of course, each spectroradiometer measurement taken in the field should be calibrated using a standard reference panel (e.g., Spectralon; Figure 11-7a). Spectroradiometer measurements made in the lab under controlled illumination conditions are valuable as long as the materials (especially vegetation samples) have been properly maintained and analyzed as soon as possible. *In situ* spectroradiometer spectral reflectance measurements can be useful in the successful transformation of hyperspectral data from radiance to scaled surface reflectance. If *in situ* spectroradiometer measurements cannot be obtained at the time of the remote sensing data collection, then it may be possible to locate some *in situ* spectra of the features of interest in established digital libraries (e.g., Baldridge, et al., 2009; ASTER Spectral Library, 2014).

If possible, a radiosonde launched at the time of remote sensing data collection can provide very valuable

information about atmospheric temperature, pressure, relative humidity, wind speed, ozone, and wind direction aloft.

Absolute Atmospheric Correction

Chapter 6 provides information about both relative and absolute atmospheric correction. This section only describes aspects of absolute atmospheric correction for analysis of hyperspectral data.

Radiative Transfer-Based Absolute Atmospheric Correction

Ideally, the analyst knows the exact nature of the atmospheric characteristics above each picture element at the time of hyperspectral data collection, e.g., barometric pressure, water vapor, amount of atmospheric molecular (Rayleigh) scattering, etc. Unfortunately, the atmosphere is variable even over relatively short geographic distances and time periods. One method of atmospheric correction uses the remote sensing–derived radiance data in very selective narrow bands to infer information about the atmospheric conditions above each pixel. This information is then used to remove atmospheric effects from each pixel. The output of radiative transfer-based atmospheric correction algorithms is scaled surface reflectance data stored in a hyperspectral image cube that has the same dimensions as the input image cube (Figure 11-5d). Several robust algorithms used to remove the effects of atmospheric attenuation from individual pixels of hyperspectral data include:

- ATmosphere REMoval (ATREM) (Boardman, 1998; Gao et al., 1999, 2009; Hatala et al., 2010; Center for the Study of Earth From Space at the University of Colorado, Boulder),
- Atmospheric CORrection Now (ACORN) (San and Suzen, 2010; ImSpec ACORN, 2014),
- Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) (Agrawal and Sarup, 2011; Exelis FLAASH, 2014),
- Quick Atmospheric Correction (QUAC) (Bernstein et al., 2008; Agrawal and Sarup, 2011; Exelis QUAC, 2014), and
- ATmospheric CORrection (ATCOR) (San and Suzen, 2010; Richter and Schlaper, 2012; ERDAS ATCOR, 2014).

Detailed characteristics associated with ACORN, FLAASH, QUAC, and ATCOR were provided in Chapter 6.

Alexander Goetz first perceived the need for developing an atmospheric correction technique using radiative transfer modeling for hyperspectral image analysis, given the limitations of the empirical approaches for surface reflectance retrievals (Gao et al., 2009). The thought was soon realized with the development of the Atmosphere Removal algorithm (ATREM) (Gao et al., 1993) for retrieving land surface reflectance spectra from hyperspectral imaging data using a theoretical modeling technique, which simulates explicitly the absorption and scattering effects of atmospheric gases and aerosols (Gao et al., 2009). The development of ATREM was a significant event in the history of imaging spectrometry.

ATREM can retrieve surface reflectance from imaging spectrometer data without the use of *in situ* spectroradiometer measurements. It atmospherically corrects the data based on water vapor absorption bands at 0.94 μm and 1.14 μm. It also determines the transmission spectra of various atmospheric gases and the scattering effects due to atmospheric molecules and aerosols. The measured radiances are divided by solar irradiance above the atmosphere to yield apparent surface reflectance. ATREM continues to be upgraded (e.g., it now uses 6S instead of 5S for modeling atmospheric scattering). ATREM is not a commercial product but has been provided to selected users that have study areas with substantial atmospheric gaseous absorption features through the Center for the Study of Earth from Space, University of Colorado at Boulder (Gao et al., 2009).

To appreciate the importance of atmospherically correcting hyperspectral data prior to data analysis, consider the application of an ATREM absolute radiative transfer-based correction to run 08 scene 05 of the Savannah River Site. An example of a raw AVIRIS loblolly pine (*Pinus taeda*) spectrum for a pixel located at 3,683,437 Northing and 437,907 Easting UTM (sample 198, line 123) is shown in Figure 11-8a. The *y*-axis measures at-sensor radiance, L (μW cm^{-2} nm^{-1} sr^{-1}). For all practical purposes, the raw radiance spectra is not very informative for Earth surface inquiries. Notice the dramatic increase in radiance in the shorter wavelength (blue) portion of the spectrum caused by increased atmospheric Rayleigh scattering. Basically, the radiance spectra do not resemble the spectra that should be associated with a healthy, photosynthesizing pine canopy.

Application of the atmospheric correction program to the raw AVIRIS data resulted in the scaled surface reflectance spectrum shown in Figure 11-8b. The scaled reflectance now appears as it should, with chlorophyll absorption bands in the blue and red, a peak in the green portion of the spectrum, increased reflectance throughout the near-infrared, and depressions surrounding the atmospheric absorption bands. Note that

Pixel-by-Pixel Atmospheric Correction of AVIRIS Data

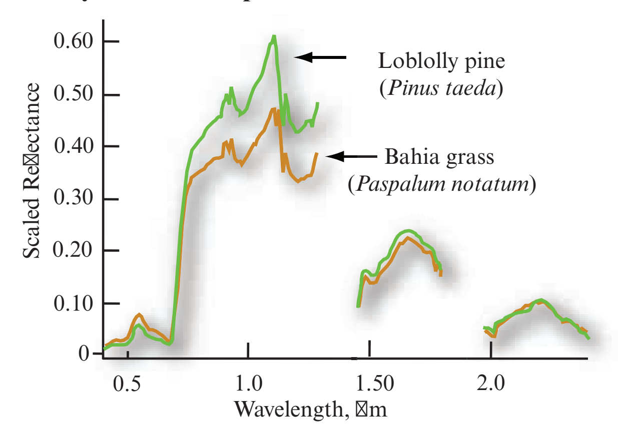
At-sensor Radiance, *L* (scaled)
5,000
4,000
3,000
2,000
1,000
0
0.5
1.0
1.50
2.0
2.5
Wavelength, μm
Loblolly pine
(*Pinus taeda*)

a. Run 08 scene 05,
original AVIRIS data.

Scaled Reflectance
0.60
0.50
0.40
0.30
0.20
0.10
0
0.5
1.0
1.50
2.0
2.5
Wavelength, μm
Loblolly pine
(*Pinus taeda*)

b. Run 08 scene 05,
atmospherically corrected using ATREM.

Band-by-Band Atmospheric Correction of AVIRIS Data



c. Run 08 scene 05,
ATREM- and EFFORT-corrected AVIRIS data.

FIGURE 11–8 a) An example of an original AVIRIS vegetation spectral profile for a single pixel of loblolly pine located at 3,683,437 Northing and 437,907 Easting in the Universal Transverse Mercator (UTM) projection. The *y*-axis is in units of radiance (*L*). b) The results of applying an absolute atmospheric correction algorithm (ATREM) to the original AVIRIS data. This is the scaled surface reflectance data for the loblolly pine pixel. c) The results of applying the Empirical Flat Field Optimal Reflectance Transformation (EFFORT) to the atmospherically-corrected loblolly pine data in (b). This resulted in a mild band-by-band, entire-scene correction. Some of the artifacts of the atmospheric correction have been partially removed in the EFFORT-corrected spectra. The scaled surface reflectance for a pixel of Bahia grass (*Paspalum notatum*) in the same scene is also plotted.

discontinuities in the spectra represent atmospheric absorption bands that are not plotted.

The quality of the atmospheric correction and radiometric calibration to scaled surface reflectance is very important. If the atmospheric correction and radiometric calibration are poor, it will be difficult for the analyst to compare remote sensing–derived spectra with *in situ*–derived spectra stored in spectral libraries.

Band-by-Band Spectral Polishing Notice in Figure 11-8b that there is noise in the loblolly pine (*Pinus taeda*) spectrum even after the data were atmospherically corrected on a pixel-by-pixel basis. This is because there is still cumulative error in the hyperspectral dataset due to sensor system anomalies and the limited accuracy of the standards, measurements, and models used and calibrations performed along the signal processing chain. This cumulative error can be as much as a few percent in each band of data.

There are additional techniques that can be used to remove some of this cumulative error as discussed in Chapter 6. Normally this involves deriving a relatively

mild linear correction that is applied to the entire band and the entire scene at one time (i.e., it is not a pixel-by-pixel correction). The methods "polish" out some of this error in an attempt to improve the accuracy of the scaled surface reflectance data (Boardman, 1997).

ENVI, ACORN, and ATCOR all have spectral polishing (artifact suppression) capability. The spectral polishing modules require atmospherically-corrected data as input. In addition, the polishing programs can also incorporate *in situ* spectroradiometer spectra for those materials most commonly found in the scene.

For example, ENVI's Empirical Flat Field Optimal Reflectance Transformation (EFFORT) program performs spectral polishing (Hatala et al., 2010; Exelis ENVI, 2014; Exelis FLAASH, 2014). EFFORT was run using the atmospherically-corrected data previously discussed. In addition, two *in situ* spectroradiometer sweet gum (*Liquidambar styraciflua*) spectra (one representing an average of multiple individual spectra, and one unaveraged spectrum) and two loblolly pine spectra were also input. This set of four (4) spectra was applied to the atmospherically-corrected bands. An EFFORT treatment was also applied using a laboratory-derived Bahia grass spectrum, but it was applied to only a limited number of visible and near-infrared channels due to its limited spectral range (400 – 800 nm). Bahia grass was judged to be important because it is grown on many of the clay-caps. Loblolly pine and sweet gum are important because they are the vegetation types often impacted downslope of clay-capped SRS waste sites.

These *in situ* spectral reflectance measurements are sometimes referred to as "reality boost spectra." Figure 11-8c depicts the same loblolly pine pixel in the hyperspectral dataset after it was spectrally polished using EFFORT. Notice that some artifacts in the atmospheric correction are removed and that the vegetation spectrum is improved. The spectral reflectance curve for a nearby pixel of Bahia grass (*Paspalum notatum*) is also shown in Figure 11-8c for comparison. Note the dramatic differences in the reflectance characteristics between the loblolly pine and Bahia grass, especially in the visible and near-infrared portion of the spectrum.

Before running programs like EFFORT, it is important that the user avoid wavelength ranges that contain noise such as the 1.4 μm and 1.9 μm water vapor absorption bands. Also, the user should experiment with the order of the polynomial that is fit to the data. The lower the order of the polynomial, the more error suppression, but the greater the potential loss of valuable spectral information. A higher-order polynomial will produce a spectrum that fits the data better but also may fit some error features. The example was based on a 10th-order polynomial. Programs like EFFORT perform a very mild adjustment to the atmospherically-corrected data. An additional method of improving the EFFORT spectral polishing is to carefully subset out only those areas of greatest importance and use "reality boost" spectra specific to that area of interest. Adler-Golden et al. (2002) provide an algorithm for removing shadows in spectral imagery. Pignatti et al. (2009) and Hatala et al. (2010) used ATREM and EFFORT polishing to prepare hyperspectral imagery for extraction of National Park ecosystem information in Italy and Wyoming.

Absolute Atmospheric Correction using Empirical Line Calibration

As discussed in Chapter 6, it is also possible to use *in situ* spectroradiometric data to atmospherically correct hyperspectral data using empirical line calibration. This method applies a single adjustment to all the pixels in a single band based on a model that is derived from the regression of *in situ* spectroradiometer measurements at specific locations (e.g., deep nonturbid water bodies, asphalt parking lots, bare soil, concrete) with radiance measurements extracted from the hyperspectral data at the same geographic locations. The *in situ* spectral measurements can also come from established spectra libraries such as the comprehensive ASTER spectral library (Baldridge et al., 2009; ASTER Spectral Library, 2014). An empirical line calibration example was provided in Chapter 6.



Geometric Correction of Hyperspectral Remote Sensor Data

It is important to geometrically correct the remote sensor data to a known datum and map projection at some point in the hyperspectral data analysis chain. Some image analysts prefer to perform all the information extraction on unrectified, atmospherically corrected hyperspectral data, and then at the very end geometrically correct the derived information for distribution to the public. Other scientists geometrically correct the hyperspectral data early in the process because they want to relate *in situ* spectroradiometer spectral reflectance measurements at known *x, y* locations in the field with the same locations in the hyperspectral data. It is also useful to geometrically rectify the hyperspectral data early in the processing chain if field collected data are to be used to train a classifier. For example, the GPS *x, y* coordinates of a lake polygon inventoried on the ground might be used to extract training statistics or identify candidate water end-member pixels in a rectified hyperspectral dataset. This

operation cannot be performed if the hyperspectral data are unrectified.

The geometric rectification techniques described in Chapter 7 may be applied to hyperspectral data. The process just takes relatively longer because there are more bands to process. Most scientists use nearest-neighbor resampling when geometrically correcting hyperspectral data. However, *n*-term polynomial and rubber-sheet image warping techniques sometimes do not yield satisfactory results when applied to aircraft-derived imaging spectrometer data. In such cases, onboard engineering and navigation data are also used. Properly configured aerial platforms are usually equipped with onboard GPS and inertial navigation system (INS) technology, which provide data that can be applied to the geometric correction problem. Jensen et al. (2011) describe the use of point and *linear* features in hyperspectral imagery that can be used to rectify the remote sensor data.



Reducing the Dimensionality of Hyperspectral Data

The number of spectral bands associated with a remote sensing system is referred to as its *data dimensionality*. Many remote sensing systems have relatively low data dimensionality because they collect information in relatively few regions of the electromagnetic spectrum. For example, the SPOT 1–3 HRV sensors recorded spectral data in three coarse bands. The Landsat 7 ETM$^+$ records data in seven bands. *Hyperspectral* remote sensing systems such as AVIRIS and MODIS obtain data in 224 and 36 bands, respectively. *Ultraspectral* remote sensing systems collect data in many hundreds of bands or in bands with very small bandwidths (Meigs et al., 2008). Ultraspectral sensing may eventually enable the identification of very fine molecular absorption or emission line features such as hydrocarbon gases (e.g., benzene).

As one might expect, data dimensionality is one of the most important issues when analyzing remote sensor data. The greater the number of bands in a dataset (i.e., its dimensionality), the more pixels that must be stored and processed by the digital image processing system. Storage and processing consume valuable resources. Therefore, significant attention has been given to developing methods to reduce the dimensionality of hyperspectral data while retaining the information content inherent in the imagery.

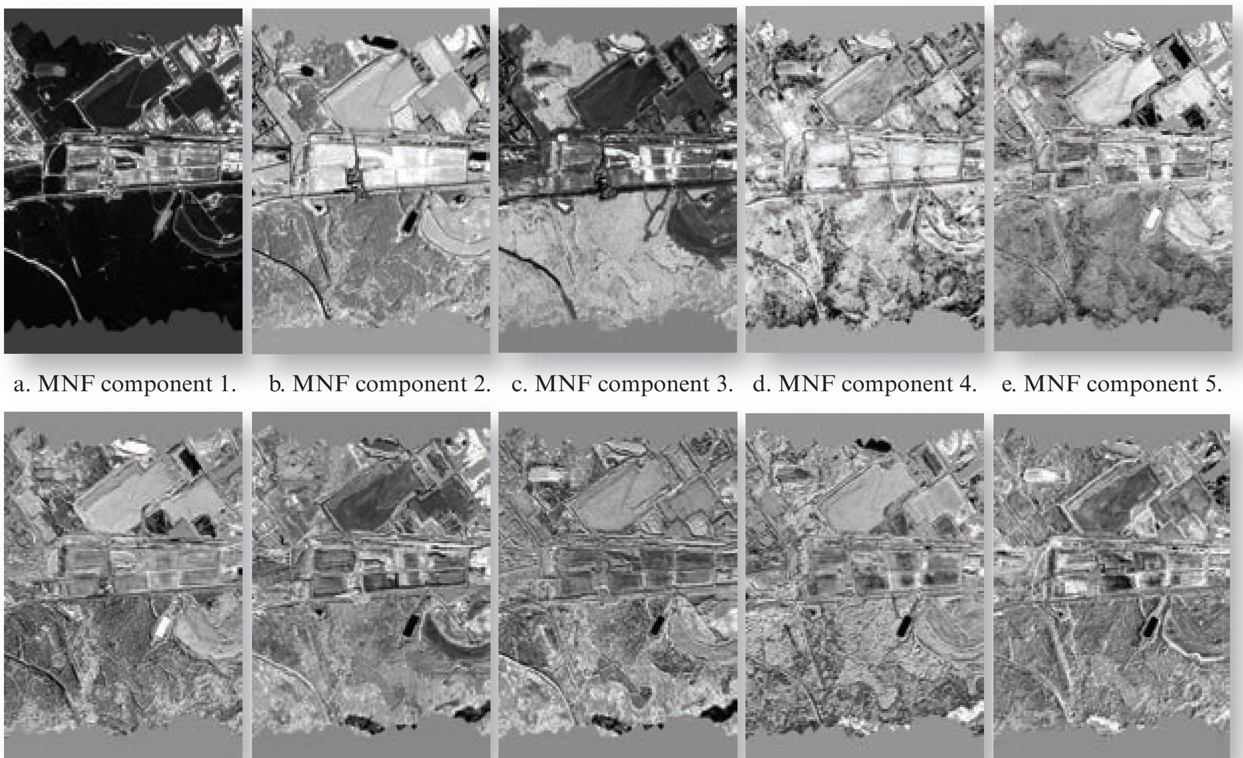
Statistical measures such as the optimum index factor (OIF), transformed divergence, and principal components analysis have been used for decades to reduce the dimensionality of multispectral data. These methods are described in Chapters 7, 8, and 9. Unfortunately, these methods generally are not sufficient for reducing the dimensionality of hyperspectral data.

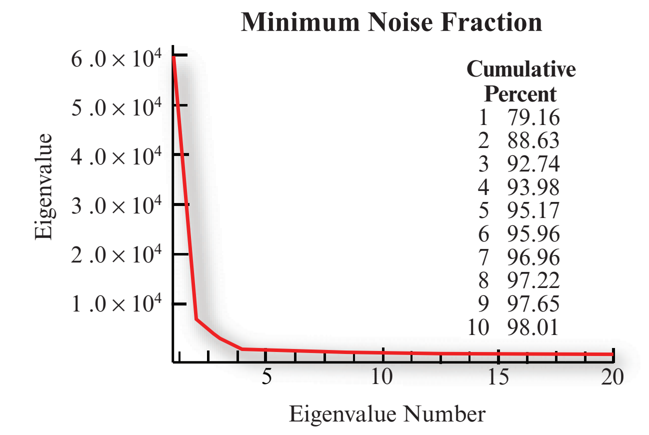
Hyperspectral data contain a tremendous amount of redundant spectral information. This is not surprising when one considers that the individual channels often have a nominal bandwidth of only 10 nm. Thus, in the spectral region from 820 to 920 nm we can expect to find 10 bands measuring the amount of near-infrared radiant flux exiting the Earth's surface. While there are certainly subtle differences in the amount of radiant flux recorded in each of these bands, there will probably be a significant amount of redundant spectral information. Statistical analysis usually reveals that many of these 10 bands are highly correlated. Therefore, one can use statistical methods to a) delete some of the unwanted redundant bands, or b) transform the data so that the information content is preserved while reducing the dimensionality of the dataset. In addition, it is hoped that the selected data dimensionality reduction method will also remove some of the noise in the hyperspectral dataset.

Minimum Noise Fraction (MNF) Transformation

A useful algorithm for reducing the dimensionality of hyperspectral data and minimizing the noise in the imagery is the minimum noise fraction (MNF) transformation (e.g., Belluco et al., 2006; Small, 2012). The MNF is used to determine the true or inherent dimensionality of the hyperspectral data, to identify and segregate noise in the data, and to reduce the computation requirements of further hyperspectral processing by collapsing the useful information into a much smaller set of MNF images (Boardman and Kruse, 1994). The MNF transformation applies two cascaded principal components analyses (Chen et al., 2003). The first transformation decorrelates and rescales noise in the data. This results in transformed data in which the noise has unit variance and no band-to-band correlation. A second principal components analysis results in the creation of a) coherent MNF eigenimages that contain useful information, and b) noise-dominated MNF eigenimages. Basically, the noise in the hyperspectral dataset is separated from the useful information. This is very important because subsequent hyperspectral data analysis procedures function best when the hyperspectral data contain coherent, useful information and very little noise.

Both the eigenvalues and the output MNF eigenimages are used to determine the *true* dimensionality of the data. Application of the MNF transformation often results in a significant decrease in the number of bands that must be processed during subsequent endmember





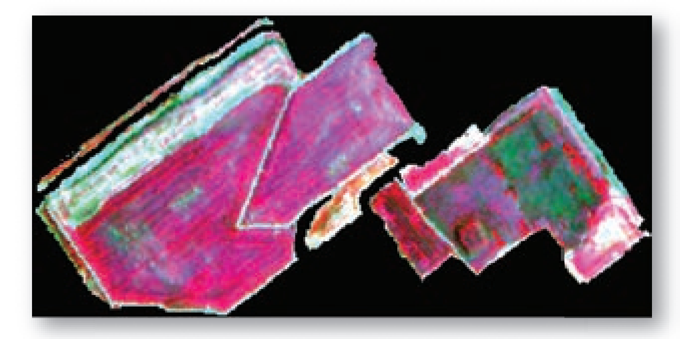
k. Plot of eigenvalues associated with 19 minimum noise fraction (MNF) bands and the cumulative amount of variance accounted for by the first 10 eigenimages.

FIGURE 11–9 a–j) The first 10 eigenimages (MNF bands) extracted from run 08 scene 05. k) A plot of the eigenvalues associated with the first 19 MNF eigenimages. Note that as the MNF eigenimage number increments, the eigenvalue and the coherent information in the eigenimages decrease. However, there is still a degree of spatial coherency in MNF band 10. Original AVIRIS imagery courtesy of NASA.

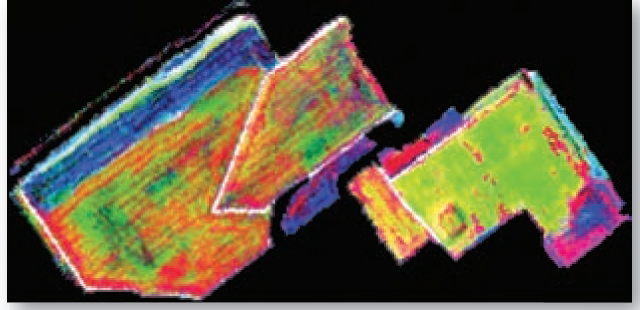
analysis. In the case of AVIRIS data, the process can result in a reduction from 224 bands to <20 useful MNF bands. This number of bands is actually quite manageable and exhibits the dimensionality of typical multispectral datasets. Therefore, some scientists have input the MNF bands that describe most of the variance directly into traditional classification algorithms (e.g., maximum likelihood) (Underwood et al., 2003). In the SRS example, the bands dominated by atmospheric noise were discarded prior to running the MNF transformation.

MNF output bands that contain useful image information typically have an eigenvalue an order of magnitude greater than those that contain mostly noise. An analyst can determine the information content of individual MNF eigenimages by a) displaying them on the computer screen and visually analyzing them, or b) plotting their eigenvalues or the amount of variance explained by each eigenvalue. For example, an MNF transformation was applied to AVIRIS run 08 scene 05. Figure 11-9 depicts the first 10 MNF eigenimages. The first 19 MNF bands could be selected for further

Masked AVIRIS Hyperspectral Data of the Clay-capped Mixed Waste Management Facility



a. False-color composite (RGB = bands 50, 30, 20).



b. MNF false-color composite (RGB = MNF 3, 2, 1).

FIGURE 11–10 a–b) Only the clay-capped hazardous waste sites associated with the Mixed Waste Management Facility and some internal roads are included in the July 26, 1999, AVIRIS imagery. All other land cover was masked from further analysis. Original AVIRIS imagery courtesy of NASA.

analysis based on a dual inspection of the spatial coherency of the eigenimages and the corresponding eigenvalue plot. Generally, the more spatially coherent the image, the less noise and greater the information content. A graph of the eigenvalues by band reveals that the first 10 eigenimages contain most of the valuable information. MNF eigenimages with values close to 1 contain mostly noise. MNF bands with eigenvalues >1 account for most of the variance in the hyperspectral dataset.

Care should be exercised in the selection of the MNF bands to keep for subsequent hyperspectral image processing. For example, only the first 19 MNF bands associated with run 08 scene 05 were deemed useful for subsequent hyperspectral endmember analysis of the Savannah River Site study area, though a smaller subset could be justified. The number of MNF transform eigenimages passed to subsequent analytical steps (and hence, the inherent dimensionality of the dataset) will determine the number of endmembers, or unique materials that can be defined (see next section).

The aforementioned example was based on an entire scene of AVIRIS data. It is often useful to mask and/or subset the scene according to the most important features of interest. For example, consider Figure 11-10 where only the clay-capped hazardous waste sites associated with the Mixed Waste Management Facility are retained after masking. Figure 11-10a is a masked false-color composite of three AVIRIS bands (RGB = 50, 30, 20). Figure 11-10b is a masked false-color composite of MNF bands (RGB = 3, 2, 1). A binary mask was applied to the image such that features not of interest, including null data values, were excluded from further analysis. The MNF transform was then applied to the masked dataset, and the result was passed to the next step in the hyperspectral image processing chain.



Endmember Determination: Locating the Spectrally Purest Pixels

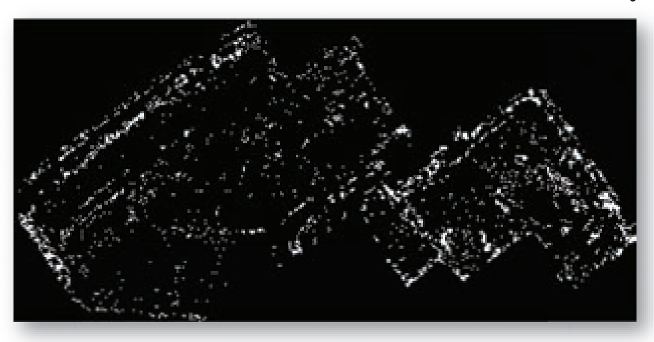
The primary goal of most hyperspectral analyses is to remotely identify the physical or chemical properties of materials found within the IFOV of the sensor system. **Spectral mixture analysis** (SMA) and its variants are powerful algorithms that can be used to determine the relative proportions of pure endmember spectra in a pixel. **Endmember** spectra are target land-cover classes believed to be present in the data that are unmixed or pure (Roth et al., 2012). Potential endmember spectra can be collected from spectroradiometer measurements collected in the lab, in the field, simulated using radiative transfer modeling and/or directly from the satellite or airborne remote sensor data (Roth et al., 2012).

Most of the time the spectral reflectance entering the IFOV of the sensor system at the instant of data collection is a function of the radiant flux from a variety of endmember materials. If we can identify the spectral characteristics of the endmember materials, then it may be possible to identify pixels that contain varying *proportions* of these materials. There are several methods of identifying the most spectrally pure pixels in a multispectral or hyperspectral scene, including:

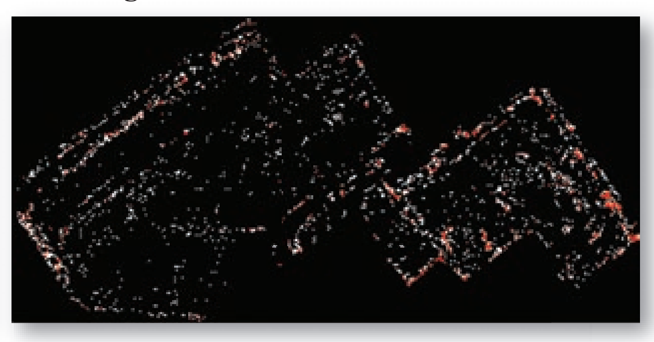
- pixel purity index, and
- *n*-dimensional visualization of endmembers in feature space.

In the Savannah River Site study, the goal was to determine if there were any vegetation stress endmembers associated with Bahia or Centipede grass growing on flat clay-capped hazardous waste sites.

Pixel Purity Index Images



a. Original.



b. Threshold at 10 hits highlighted in red.

FIGURE 11–11 Pixel purity index (PPI) image of a portion of run 08, scene 05 derived from masked minimum noise fraction AVIRIS data. The pixels in (a) denote the original output from the PPI analysis. The pixels highlighted in red in (b) represent a subset of those pixels selected using a threshold of 10 hits. Original AVIRIS imagery courtesy of NASA.

Pixel Purity Index Mapping

It is sometimes difficult to identify pure endmembers in a remote sensing dataset because very few pixels contain just one type of biophysical material. Therefore, a rigorous mathematical method of determining the most spectrally pure pixels is to repeatedly project *n*-dimensional scatterplots of the clean minimum noise fraction images (e.g., not just 3 bands as in the previous example but *all* useful MNF bands) onto a random unit vector. Each time the spectral data are projected, the most extreme pixels in each projection are noted (Research Systems, Inc., 2004; Exelis ENVI, 2014). By keeping track of which pixels in the scene are repeatedly identified as "extreme" pixels, it is possible to create a *pixel purity index* (PPI) image. Basically, the higher the pixel value in the pixel purity index image, the greater the number of times it was judged to be a spectrally extreme pixel (e.g., a pure water, pure concrete, or pure vegetation endmember). It is important to include as input to the PPI calculation only MNF images that contain spectrally valuable information.

A pixel purity index image derived from an analysis of masked minimum-noise-fraction images is shown in Figure 11-11a. Note that there are not a tremendous number of relatively pure pixels in the scene. In fact, only 4,155 pixels in the masked clay-cap image were identified as being pure to some degree. Furthermore, a threshold is typically used to create a subset of the most pure PPI pixels to be used in subsequent analyses (Figure 11-11b). In this case, only 1,207 of the 4,155 pixels were selected for further processing after 70,000 PPI iterations using a conservative threshold value of 10 "hits."

A PPI image simply identifies the location of the most pure pixels (i.e., endmember candidates). Unfortunately, it is difficult to label the types of endmembers by just viewing the two-dimensional pixel purity index image. Therefore, it is common to use an *n*-dimensional visualization technique.

n-Dimensional Endmember Visualization

To identify and label the most spectrally pure endmembers in the two-dimensional pixel purity index image, it is necessary to systematically view where the relatively few pure vegetation pixels reside in *n*-dimensional feature space or in the image itself. For example, Figure 11-12 depicts a display of the most spectrally pure pixels found in the PPI image shown in Figure 11-10b. By using more than two of the MNF bands at a time it is possible to interactively view and rotate the endmembers in *n*-dimensional spectral space on the computer screen. The analyst attempts to locate the corners of the data cloud in *n*-dimensional space as the cloud rotates. The purest pixels lie at the convex corners of the data cloud. Clusters can be defined manually, or an algorithm can be used to precluster the data, and the endmembers can be subsequently refined manually. In this context, *n* equals the number of MNF bands, and given the dimensionality of the simplex formed by the cluster means, the number of endmembers that can be defined is equal to $n + 1$. Thus, there are theoretically $n + 1$ endmembers if the data are truly *n*-dimensional (Boardman, 1993).

By comparing the actual spectra of an endmember found within the *n*-dimensional display with where it is actually located in *x*, *y* space in the MNF or surface reflectance image, it is possible to label the pure pixel as a specific type of endmember, e.g., water, vegetation, or bare soil. Endmembers that are extracted in this manner may then be used to perform spectral matching or

Three-Dimensional Visualization of Five Endmembers

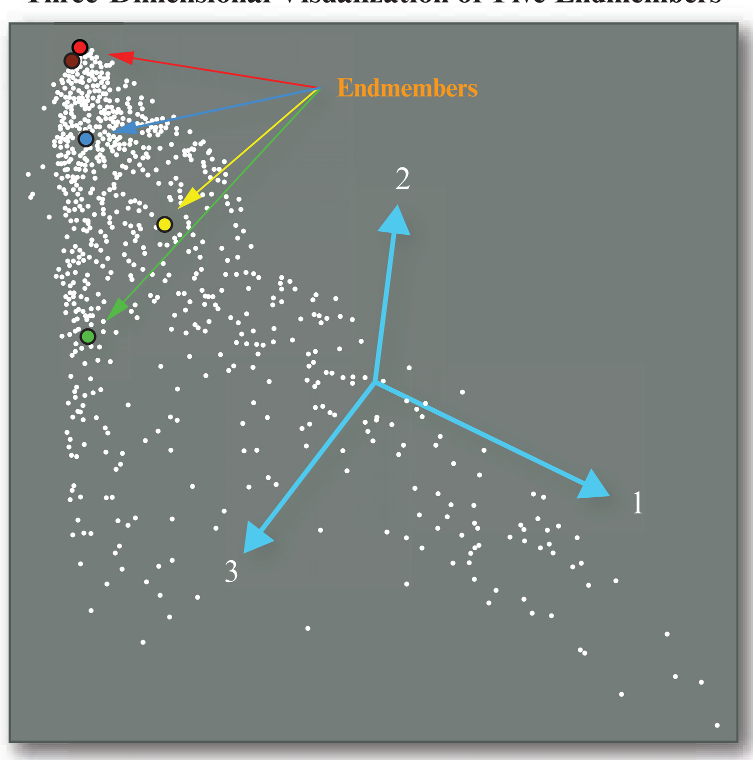


FIGURE 11–12 *N*-dimensional visualization of four potential image-derived vegetation endmembers and one shade endmember derived from the pixel purity index (PPI) analysis. Endmembers associated with other materials (e.g., asphalt, concrete) are not highlighted. Note that some of the highlighted pixels are not immediately recognizable as endmembers. Data cloud rotations were necessary to discern all the endmember regions of interest.

classification. More automated methods of endmember determination are available.

To demonstrate these methods, let us examine the clay-capped hazardous waste sites in the run 08 scene 05 dataset to identify and label the following spectra:

- healthy Bahia grass, and
- potentially stressed Bahia grass.

Masked MNF images were analyzed to produce the pixel purity index image (and spectral profiles of the most pure pixels) previously discussed (Figure 11-9). These data were then displayed as data clouds in an *n*-dimensional feature space (Figure 11-12). Spectra (scaled reflectance values) for one healthy vegetation endmember (37) and three potentially stressed vegetation endmembers (15, 25, and 36) over the wavelength interval from 0.4 to 2.43 μm are superimposed in Figure 11-13a. The same spectra are replotted in the visible through near-infrared region (0.4–0.9 μm) in Figure 11-13b to highlight certain reflectance characteristics. The numerical names assigned to these endmember spectra (e.g., #37) are an artifact of the iterative endmember selection process.

The clay-cap on the Mixed Waste Management Facility (MWMF) contains a near-monoculture of Bahia grass (there are a few patches of Centipede). Bahia grass endmember 37 exhibited all the characteristics of healthy vegetation, including a) strong chlorophyll *a* and *b* absorption bands in the blue (0.43 – 0.45 μm) and red (0.65 – 0.66 μm) portions of the spectrum, b) a relative lack of absorption in the wavelength interval between the two chlorophyll absorption bands, producing an increase in reflectance at approximately 0.54 μm in the green portion of the spectrum, c) a substantial increase in reflectance throughout the near-infrared portion of the spectrum (0.7 – 1.2 μm) due to scattering and reflectance in the spongy mesophyll portion of the plant, and d) typical absorption/reflectance characteristics in the middle-infrared (1.45 – 1.79 μm and 2.08 – 2.35 μm) wavelengths (Jensen, 2007).

Conversely, each potentially stressed Bahia grass endmember (15, 25, and 36) exhibited spectral reflectance characteristics that were very different from the healthy Bahia grass spectra. Relative to the healthy turfgrass spectrum, each exhibited a) a slight to moderate increase in red reflectance (endmember 15 most; 25 intermediate; 36 least); b) a slight to moderate red-edge

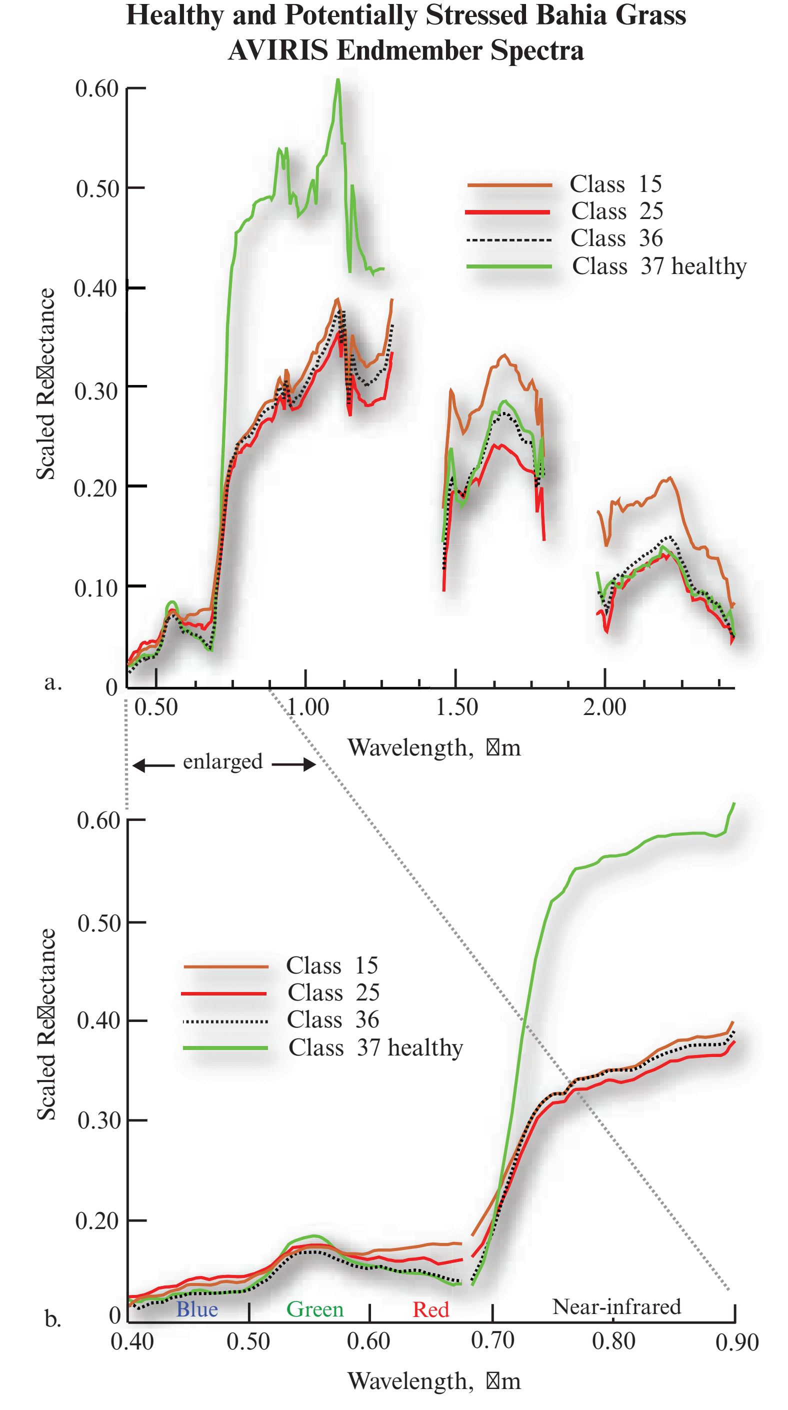


FIGURE 11–13 a) One healthy (37) and three potentially stressed Bahia grass image-derived endmember spectral profiles derived using a pixel purity index image and *n*-dimensional visualization endmember analysis. b) Only the visible through near-infrared region (0.4–0.90 μm) is plotted to highlight the increase in red reflectance, the apparent red-edge shift, and decrease in near-infrared reflectance of the potentially stressed vegetation endmembers.

shift toward shorter wavelengths in the transition region from red to near-infrared wavelengths (15 most; 25 intermediate; 36 least), and c) a significant decrease in near-infrared reflectance.

Endmember 15 consistently reflected more energy in both middle-infrared regions (1.45 – 1.79 μm and 2.08 – 2.35 μm) than the healthy Bahia grass endmember (37). Endmember 25 exhibited a substantial decrease in middle-infrared reflectance in the 1.45 – 1.79 μm region and approximately the same reflectance as the healthy vegetation in the 2.08 – 2.35 μm region. Finally, endmember 36 exhibited a decrease in middle-infrared reflectance in the 1.45 – 1.79 μm region and a slight increase in reflectance in the 2.08 – 2.35 μm region. While the identified endmembers are spectrally pure in

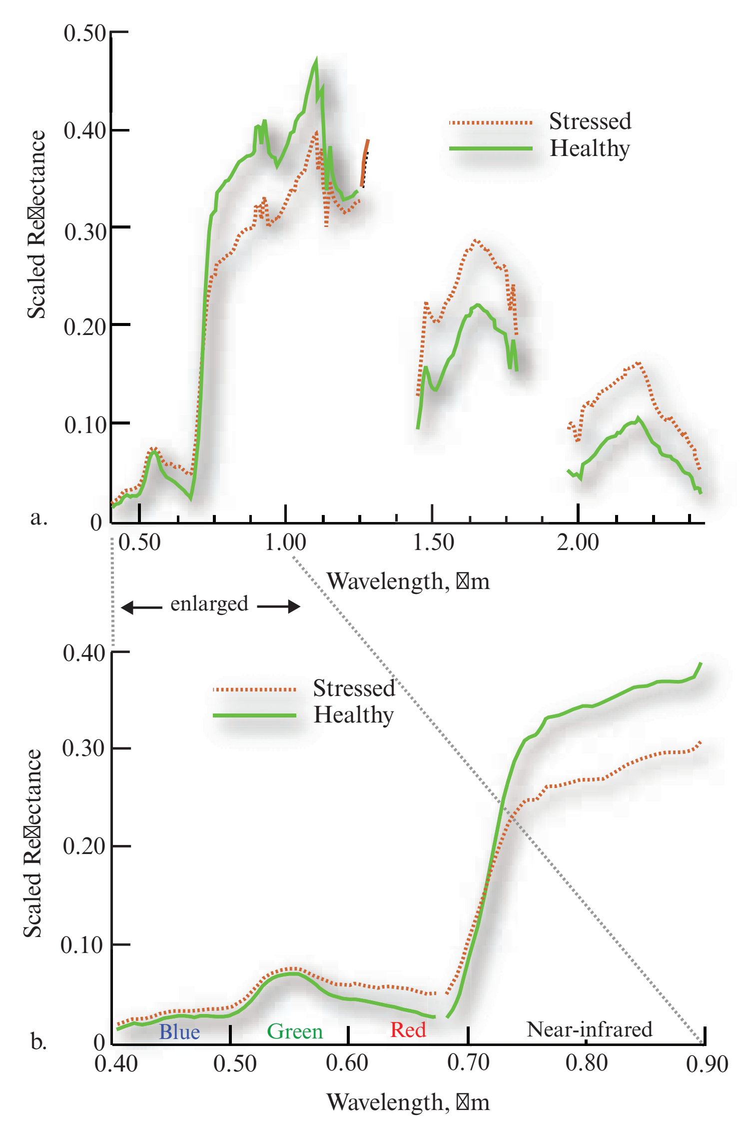


FIGURE 11–14 a) Spectra located within areas of homogeneous healthy and stressed Bahia grass. b) The visible through near-infrared region is rescaled to highlight the increase in red reflectance, the red-edge shift, and decrease in near-infrared reflectance of the potentially stressed vegetation endmembers.

theory, it is possible that albedo contributions from other materials are present in the pixels (e.g., soil background reflectance). However, the extracted pixels are the purest relative to all others in the image.

To illustrate the utility of the endmember concept further, consider the reflectance profiles obtained at two randomly selected healthy and potentially stressed areas on the Bahia grass clay-cap (Figure 11-14ab). Note that these non-endmember profiles retain the same general relationships as the image-derived endmember profiles (Figure 11-13ab).

Before proceeding, one should ask, "How can we be sure that the Bahia grass is actually undergoing some stress?" Fortunately, we have some laboratory Bahia grass spectra that depict how healthy and stressed Bahia grass reflect electromagnetic energy in a controlled environment. For example, Figure 11-15 depicts the spectra of healthy Bahia grass and Bahia grass subjected to various concentrations of copper (mg/ml) (Schuerger, 2003). Note the increase in reflectance in the red region, the red-edge shift, and the decrease in reflectance in the near-infrared region as the copper concentration is increased. Unfortunately, the laboratory spectroradiometer did not record Bahia grass spectral information in the middle-infrared region. Thus, the AVIRIS hyperspectral results only suggest that the spectra labeled "potentially stressed Bahia grass" are indicative of some type of stressing agent. It is not suggested that the stress on the MWMF is due to the absorption of copper, but rather that the stressed

Laboratory-Derived Bahia Grass Spectra

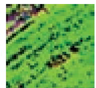
0 mg/ml Cu
0.5 mg/ml Cu
08 mg/ml Cu
25 mg/ml Cu
50 mg/ml Cu
100 mg/ml Cu

Mean Reflectance (n = 10)

Wavelength, μm

FIGURE 11–15 Laboratory spectral reflectance of healthy and copper-stressed Bahia grass in the region from 0.4 to 0.9 μm (Schuerger, unpublished data).

Bahia grass spectra simply exhibit stressed behavior throughout the visible and near-infrared portions of the spectrum (it could be completely different in the middle-infrared region). The agent responsible for the potential vegetation stress may be a) an overabundance of water, resulting in a decrease in middle-infrared reflectance, as with endmembers 25 and 36; b) too little water, resulting in an increase in reflectance in the middle-infrared region, as with endmember 15; or c) a completely different stressing agent related to clay-cap constituents.



Mapping and Matching using Hyperspectral Data

Algorithms may be used to convert hyperspectral reflectance into biophysical information or thematic maps, including the spectral angle mapper (SAM), subpixel classification (e.g., linear spectral unmixing), spectroscopic library matching, hyperspectral indices, and derivative spectroscopy. Scientists have also applied traditional classification algorithms and machine learning techniques to hyperspectral data (e.g., Goel et al., 2003; Im et al., 2009, 2012).

Spectral Angle Mapper

Spectral angle mapping (SAM) algorithms take an atmospherically corrected unlabeled pixel (e.g., composed of an AVIRIS measurement vector of n brightness values) and compare it with reference spectra in the same n-dimensions. In the SRS study, only 175 of the original 224 AVIRIS bands were used. The reference spectra may be obtained using 1) calibrated *in situ* or laboratory-derived spectroradiometer measurements stored in an ASCII or binary spectral library, 2) theoretical calculations, and/or 3) multispectral or hyperspectral image endmember analysis procedures, as previously discussed. The SAM algorithm compares the angle (α) between the reference spectrum (r) and the hyperspectral image pixel measurement vector (t) in n-dimensions (bands) (Figure 11-16) and assigns it to the reference spectrum class that yields the smallest angle (Research Systems, Inc., 2004; Belluco et al., 2006; Exelis ENVI, 2014):

$$\alpha = \cos^{-1}\left(\frac{\sum_{i=1}^{n} t_i r_i}{\left(\sum_{i=1}^{n} t_i^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{n} r_i^2\right)^{\frac{1}{2}}}\right) \quad (11.1)$$

The simple two-band example in Figure 11-16 suggests that unknown material t has reflectance spectra that are more similar to reference spectrum r than the spectrum for material k, i.e., the angle (α) in radians between r and t is smaller. Basically, for each reference spectrum r (this could be an *in situ*–derived endmember or a remote sensing–derived endmember, as with

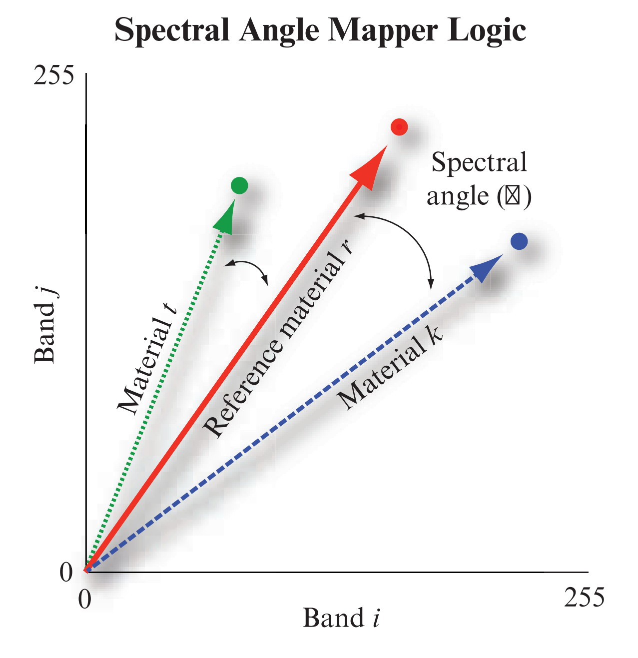


FIGURE 11–16 The logic associated with the Spectral Angle Mapper (SAM).

the four endmembers previously discussed) a spectral angle (α) is computed for each unknown image spectrum (pixel) in the dataset. This angle becomes the pixel value in the output SAM image with one output image for each reference spectrum. This creates a new SAM datacube with the number of bands equal to the number of reference spectra used in the mapping.

The SAM algorithm was applied to the previously described endmembers. Thematic maps of the spatial distribution of healthy (derived from endmember 37) and potentially stressed Bahia grass (derived from endmembers 15, 25, and 36) are shown in Figure 11-17. The darker the pixel (i.e., black, blue) the smaller the SAM angle and the closer the match. Thus, pixels in Figure 11-17a that had spectra very similar to that of the healthy Bahia grass endmember spectra (37) yielded very small angles and were color-coded as being black or dark blue. Note that the spatial distribution of the healthy vegetation derived from endmember 37 is almost the inverse of the maps derived from the potential stress endmembers (15, 25, and 36) in Figure 11-17b-d.

It is possible to convert the output from the spectral angle mapper subroutine into a hard (crisp) classification map. For example, Figure 11-18a depicts a hard classification map derived using just the four endmembers previously discussed, one class for each endmember. A threshold angle value of 0.1 radian was used. Brown areas represent healthy vegetation (based on endmember 37) and the other three classes represent potentially stressed vegetation.

The clay-capped hazardous waste sites ideally have a relatively homogenous cover of Bahia grass. Maps such as this suggest that on this date the Bahia grass may not have been uniformly distributed and that certain locations within the site may have stressed Bahia grass. This stress could be the result of a number of factors. The display provides valuable spatial information that can be of significant value to the personnel responsible for clay-cap maintenance. It can help them to focus their attention on certain locations on the clay-cap to ensure its continued integrity.

It is also possible to use the laboratory-derived Bahia grass reference spectra as endmember spectra to be input to SAM. A potentially stressed Bahia grass classification map derived from averaged laboratory spectra with a 0.5 mg/ml copper treatment is shown in Figure 11-18b. The darker the pixel (blue, black, etc.), the smaller the angle and the closer the match. It visually correlates well with the results obtained from the remote sensing–derived spectra.

Subpixel Classification, Linear Spectral Unmixing or Spectral Mixture Analysis

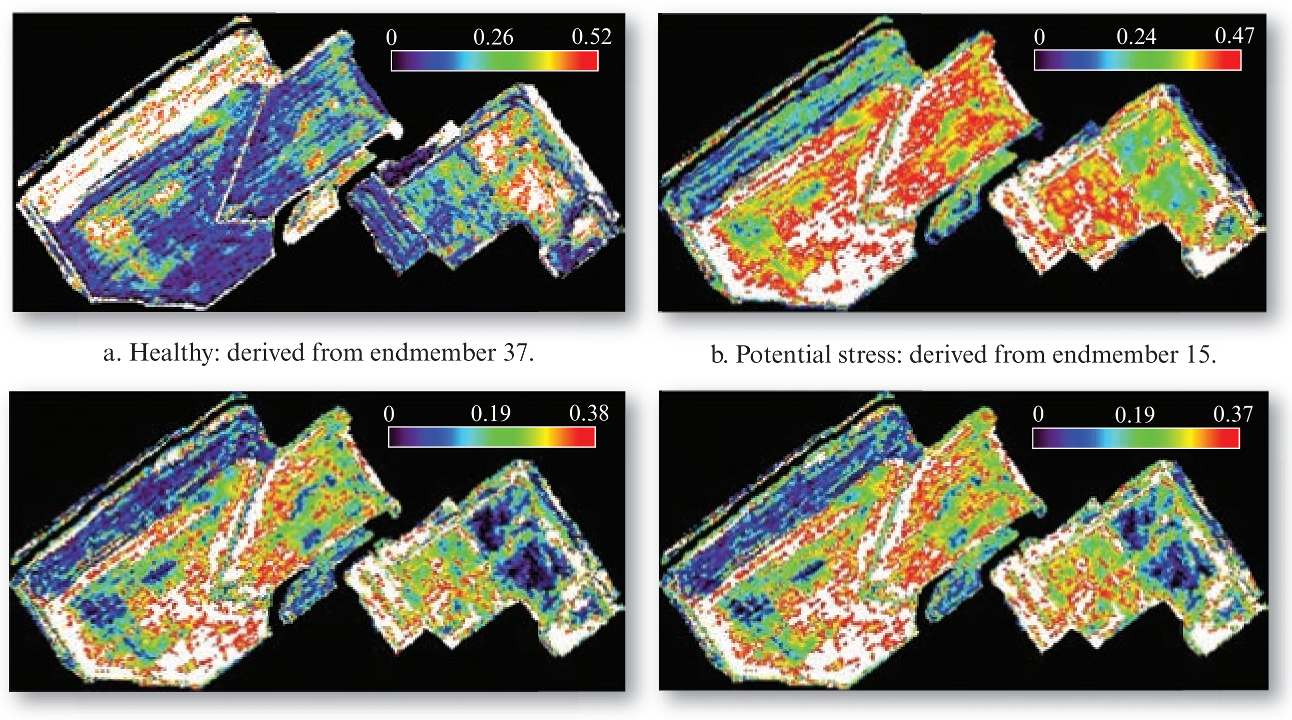
The energy recorded by a remote sensing detector is a function of the amount of energy reflected or emitted by the materials within the instantaneous-field-of-view (IFOV). For example, consider the single pixel shown in Figure 11-19a, which is composed of 50% water, 25% bare soil, and 25% vegetation. Should this pixel be classified as water because its dominant constituent is water? Wouldn't it be more useful to determine and report the actual proportion (or abundance) of the pure endmember (class) materials within the pixel? This process is commonly referred to as *subpixel classification, linear spectral unmixing,* or *spectral mixture analysis* (Okin et al., 2001; Jensen et al., 2009; Somers et al., 2011; Thenkabail et al., 2011; Prost, 2013).

Spectral mixture analysis (SMA) can be used to decompose the measured spectral reflectance within the IFOV of a single pixel as a mixture of a fixed set of endmembers. The mixing can be modeled as the sum of endmember reflectance multiplied by their fraction contribution to a best-fit mixed spectrum. Linear SMA is based on the equation (Roth et al., 2012):

$$\rho'_\lambda = \sum_{i=1}^{n} (\rho_{i\lambda} \times f_i) + \varepsilon_\lambda \qquad (11.2)$$

where ρ'_λ is the reflectance at wavelength, λ; $\rho_{i\lambda}$ is the spectral reflectance of endmember i, and f_i is the fractional cover of endmember i. The total number of endmembers in the model is n and the model error is

Classification Maps of Healthy and Potentially Stressed Vegetation Created Using Image-Derived Endmembers and a Spectral Angle Mapper Algorithm



a. Healthy: derived from endmember 37.

b. Potential stress: derived from endmember 15.

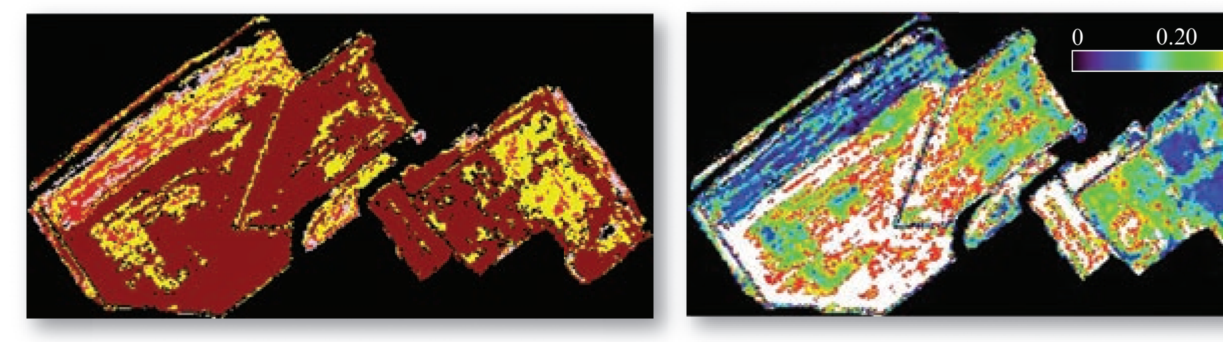
c. Potential stress: derived from endmember 25.

d. Potential stress: derived from endmember 36.

FIGURE 11–17 a) Healthy Bahia grass classification map derived from July 26, 1999, AVIRIS hyperspectral data and spectral angle mapper (SAM) analysis of endmember 37. b–d) Potentially stressed Bahia grass classification map derived from AVIRIS hyperspectral data and SAM analysis of endmembers 15, 25, and 36. In all cases, the darker the pixel (blue, black, etc.), the smaller the angle and the closer the match. Original AVIRIS imagery courtesy of NASA.

Hardened Classification Map of Healthy and Potentially Stressed Vegetation

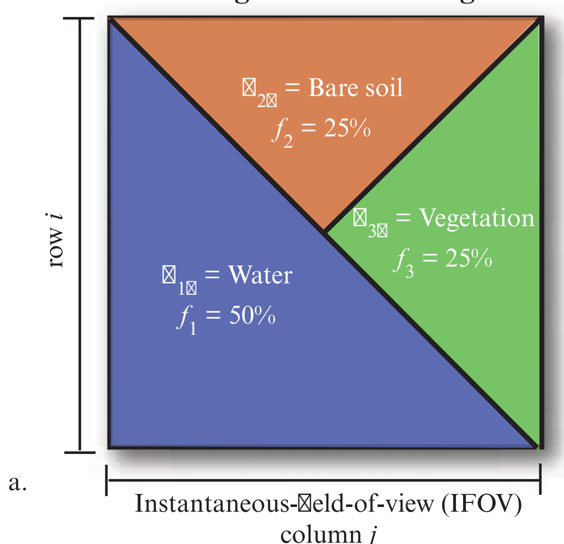
Classification Map of Potentially Stressed Vegetation Created Using Laboratory–Derived Endmembers and the Spectral Angle Mapper Algorithm



a. Hardened Spectral Angle Mapper classification map.

b. Laboratory-derived endmember classification map.

FIGURE 11–18 a) A hardened classification map derived from July 26, 1999, AVIRIS hyperspectral data and spectral angle mapper (SAM) analysis of four endmembers. Brown areas represent healthy Bahia grass. White, red, and yellow depict areas with vegetation stress-related characteristics. b) Potentially stressed Bahia grass classification map derived from laboratory spectra with a 0.5 mg/ml copper treatment. The darker the pixel (blue, black, etc.), the smaller the angle and the closer the match. Original AVIRIS imagery courtesy of NASA.



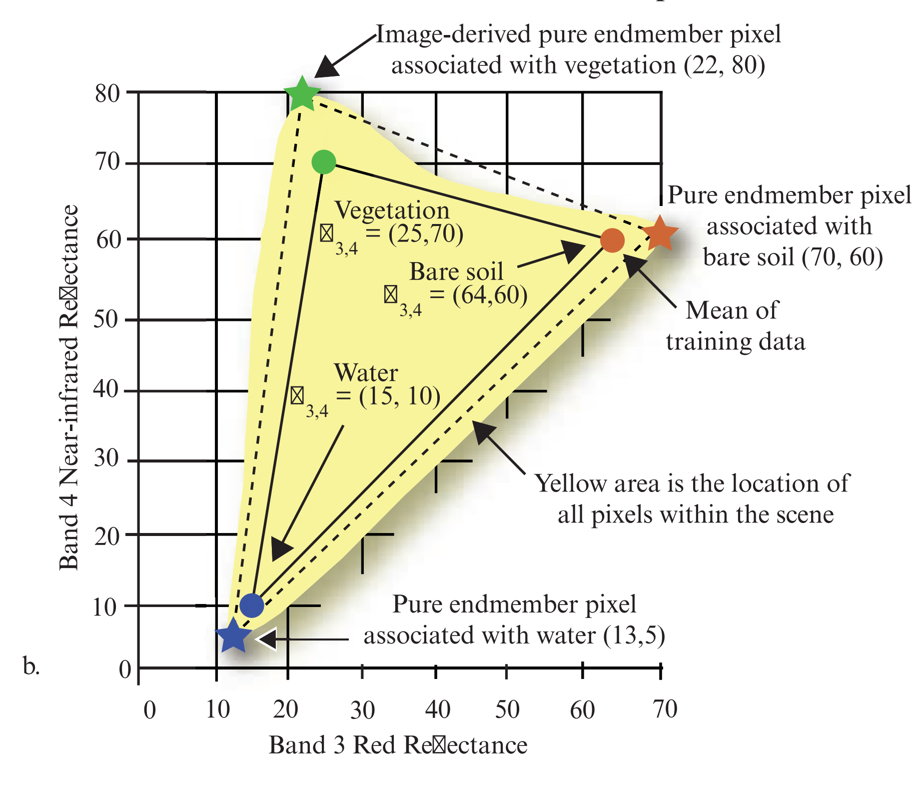


FIGURE 11–19 a) Linear mixture modeling associated with a single pixel consisting of water, vegetation, and bare soil. b) The location of hypothetical mean vectors and endmembers in two-dimensional feature space (red and near-infrared).

ε_λ. The accuracy of the SMA model fit is determined by computing the root mean square error (RMSE) (Roth et al., 2012):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{b=1}^{k} (\varepsilon_\lambda)^2}{k}} \tag{11.3}$$

where b is a band number and k is the total number of bands.

To appreciate linear spectral unmixing, let us assume for a moment that we have a remotely sensed image consisting of just two bands ($k = 2$) and that there are only three types of pure endmember materials in the

scene (n = 3): water, vegetation, and bare soil—$\rho_{1\lambda}$, $\rho_{2\lambda}$, and $\rho_{3\lambda}$, respectively. The linear mixing associated with any pixel in the scene can be described using the matrix notation:

$$[\rho'_\lambda] = [\rho_\lambda][f] + \varepsilon \qquad (11.4)$$

where $[\rho'_\lambda]$ is the k-dimensional spectral vector of the pixel under investigation, $[f]$ is the $n \times 1$ vector of n endmember fractions (e.g., 0.5, 0.25, 0.25) for the pixel and $[\rho_\lambda]$ is the $k \times n$ signature matrix, with each column containing one of the endmember spectral vectors (Schowengerdt, 2007). The relationship contains some noise, therefore ε represents the residual error. This equation assumes that we have identified all of the theoretically pure classes (endmembers) in the scene such that their proportions (f_i) will sum to 1 at each pixel and that all endmember fractions are positive.

If we assume that the data contain no noise (ε), Equation 11.4 becomes:

$$[\rho'_\lambda] = [\rho_\lambda][f] \qquad (11.5)$$

and we can use the following matrix notation to solve for $[f]$ on a pixel-by-pixel basis:

$$\begin{bmatrix} \rho'_3 \\ \rho'_4 \end{bmatrix} = \begin{bmatrix} \rho_{water3} & \rho_{veg3} & \rho_{bsoil3} \\ \rho_{water4} & \rho_{veg4} & \rho_{bsoil4} \end{bmatrix} \begin{bmatrix} f_{water} \\ f_{veg} \\ f_{bsoil} \end{bmatrix} \qquad (11.6)$$

Unfortunately, this relationship is under-determined because there are fewer bands than endmembers. Fortunately, because we know that the linear combination of all three endmembers must sum to 1 (i.e., $f_{water} + f_{vegetation} + f_{bare\ soil} = 1$), we can augment the equation to become:

$$\begin{bmatrix} \rho'_3 \\ \rho'_4 \\ 1 \end{bmatrix} = \begin{bmatrix} \rho_{water3} & \rho_{veg3} & \rho_{bsoil3} \\ \rho_{water4} & \rho_{veg4} & \rho_{bsoil4} \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} f_{water} \\ f_{veg} \\ f_{bsoil} \end{bmatrix} \qquad (11.7)$$

This equation can be inverted to solve for the exact fraction (proportion) of endmember materials found within each pixel:

$$\begin{bmatrix} f_{water} \\ f_{veg} \\ f_{bsoil} \end{bmatrix} = \begin{bmatrix} \rho_{water3} & \rho_{veg3} & \rho_{bsoil3} \\ \rho_{water4} & \rho_{veg4} & \rho_{bsoil4} \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho'_3 \\ \rho'_4 \\ 1 \end{bmatrix} \qquad (11.8)$$

TABLE 11–2 Reflectance values for the three hypothetical image-derived endmembers shown in Figure 11-19b.

Band	Water	Vegetation	Bare Soil
3	13	22	70
4	5	80	60

Please note that $[\rho_\lambda]$ is now $[\rho_\lambda]^{-1}$.

Now let us use Equation 11.5 to determine the proportion of endmember materials in pixels in a hypothetical scene. Consider the two-dimensional feature space plot shown in Figure 11-19b. This scene consists of just two bands of remote sensor data (band 3 and 4; red and near-infrared, respectively). The yellow cloud represents the co-occurrence of all pixel values in the scene in bands 3 and 4. There are no pixel values in the scene outside the yellow area convex hull. The circles are mean vectors associated with water, vegetation, and bare soil obtained through supervised training. They are present to help the reader appreciate the difference between mean vectors and endmembers and are not used in this computation. The three stars in the display represent pure endmember pixels associated with water, vegetation, and bare soil. These endmembers were derived from the image itself using pixel purity index assessment and/or n-dimensional visualization or by obtaining accurate spectroradiometer measurements on the ground of pure water, vegetation, and bare soil.

This graphic contains enough data for us to derive the information necessary to predict the proportion of each endmember in each pixel in the scene. The values found in Figure 11-19b are summarized in Table 11-2. It is possible to compute the fractions of each endmember class found in each pixel in the scene by a) placing the inverse matrix $[\rho_\lambda]^{-1}$ coefficients in Table 11-3 into Equation 11-5, and b) repeatedly placing new values of $\rho_{\lambda 3}$ and $\rho_{\lambda 4}$ in Equation 11.5 associated with each pixel in the scene. For example, if the $\rho_{\lambda 3}$ and $\rho_{\lambda 4}$ values for a single pixel were 25 and 57, respectively, then the proportion (abundance) of water, vegetation, and bare soil found within this single pixel would be 27%, 61%, and 11%, as shown below:

$$\begin{bmatrix} f_{water} \\ f_{veg} \\ f_{bsoil} \end{bmatrix} = \begin{bmatrix} -0.0053 & -0.0127 & 1.1322 \\ -0.0145 & 0.0150 & 0.1137 \\ 0.0198 & -0.0024 & -0.2460 \end{bmatrix} \begin{bmatrix} 25 \\ 57 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 0.27 \\ 0.61 \\ 0.11 \end{bmatrix} = \begin{bmatrix} -0.0053 & -0.0127 & 1.1322 \\ -0.0145 & 0.0150 & 0.1137 \\ 0.0198 & -0.0024 & -0.2460 \end{bmatrix} \begin{bmatrix} 25 \\ 57 \\ 1 \end{bmatrix}$$

TABLE 11–3 Matrices used to perform spectral unmixing.The inverse matrix $[\rho_\lambda]^{-1}$ was derived from the endmember reflectance values in $[\rho_\lambda]$.

$\left[\rho_\lambda\right]$	$\left[\rho_\lambda\right]^{-1}$
$\begin{bmatrix} 13 & 22 & 70 \\ 5 & 80 & 60 \\ 1 & 1 & 1 \end{bmatrix}$	$\begin{bmatrix} -0.0053 & -0.0127 & 1.1322 \\ -0.0145 & 0.0150 & 0.1137 \\ 0.0198 & -0.0024 & -0.2460 \end{bmatrix}$

This logic may be extended to hyperspectral imagery and the overdetermined problem, where there are more bands than endmembers. Schowengerdt (2007) describes the more complex pseudo-matrix inversion that must be adopted when analyzing hyperspectral data.

Spectral mixture analysis (SMA) assumes a pixel's spectrum is a linear combination of a finite number of spectrally distinct endmembers. Spectral mixture analysis uses the dimensionality of hyperspectral data to produce a suite of abundance (fraction) images for each endmember. Each fraction image depicts a subpixel estimate of endmember relative abundance as well as the spatial distribution of the endmember. When the endmembers include vegetation, the endmember fraction is proportional to the areal abundance of projected canopy cover (Roberts et al., 1993; Williams and Hunt, 2002). Although SMA is intuitively very appealing, scientists often find it difficult to identify all of the pure endmembers in a scene (Jensen et al., 2009).

McGwire et al. (2000) found that endmembers (average green leaf, soil, and shadow) derived from hyperspectral data in arid environments were more highly correlated with vegetation percent cover than when the hyperspectral data are processed using the traditional narrow-band and broadband vegetation indices (e.g., NDVI, SAVI, EVI). Williams and Hunt (2002) used mixture-tuned matched filtering (MTMF), a special case of subpixel analysis, to estimate leafy spurge canopy cover and map its distribution. Segl et al. (2003) used endmembers derived from hyperspectral data and linear spectral unmixing to identify urban surface cover types. Franke et al. (2009) used hierarchical multiple endmember spectral mixture analysis (MESMA) to map urban land cover. Hatala et al. (2010) used MTMF to identify forest pest and pathogen damage in the Greater Yellowstone Ecosystem. Roth et al. (2012) compared several endmember selection techniques and evaluated the tradeoffs between classification accuracy achieved, the size of the library created, and computation time.

Continuum Removal

Even when great care is given to the selection of endmembers during SMA, the classification results may not be optimal because of the spectral similarity between classes, especially for similar plant types or closely-related plant species. Youngentob et al. (2011) suggest that *continnum removal* (CR) analysis may improve class separability by emphasizing individual absorption features across a normalized spectrum. CM involves fitting a spectral reflectance curve with a convex hull (i.e., a line just touching the maximum points along the original reflectance curve). The continuum is removed by dividing reflectance values at a specific wavelength (ρ_λ) by the reflectance value of the continnum ($\rho_{c\lambda}$) at the corresponding wavelength λ. The continuum removed (CR) spectral reflectance ($\rho_{cr\lambda}$) at wavelength λ is given by (Youngentob et al., 2011; Rodger et al., 2012):

$$\rho_{cr\lambda} = \frac{\rho_\lambda}{\rho_{c\lambda}} \tag{11.9}$$

Applying continuum removal to the scaled reflectance data enhances the subtle absorption features in the spectral reflectance curves because their depth and position are not influenced by variations in albedo (Filippi and Jensen, 2007). For example, consider the reflectance curve of healthy vegetation sampled every 50 nm from 400 to 2,500 nm in Figure 11-20a and its associated continuum (convex hull). The application of Equation 11.9 produces a new vegetation reflectance curve shown in Figure 11-20b. The difference between the green reflectance maximum and the red reflectance minimum in the original reflectance curve is approximately 5–7% (Figure 11-20a). The difference between the green maximum and the red minimum reflectance is approximately 20% in the continuum-removed curve (Figure 11-20b). The emphasis of the red absorption feature (and other absorption features not discussed) in the continuum-removed data may make it possible to extract more accurate information when the data are analyzed using SMA or other techniques (e.g., NDVI).

Spectroscopic Library Matching Techniques

One of the benefits of conducting hyperspectral remote sensing research is that a detailed spectral response calibrated to percent reflectance (spectra) can be obtained for a pixel (e.g., AVIRIS measures in 224 bands from 400 to 2500 nm). The remote sensing–derived spectra can be compared with *in situ*- or laboratory-derived spectra that is stored in a special spectral library. The laboratory-derived spectra are usually considered more accurate because they have been made under controlled illumination conditions, and the atmosphere is not a factor. Laboratory-derived spectra may be found

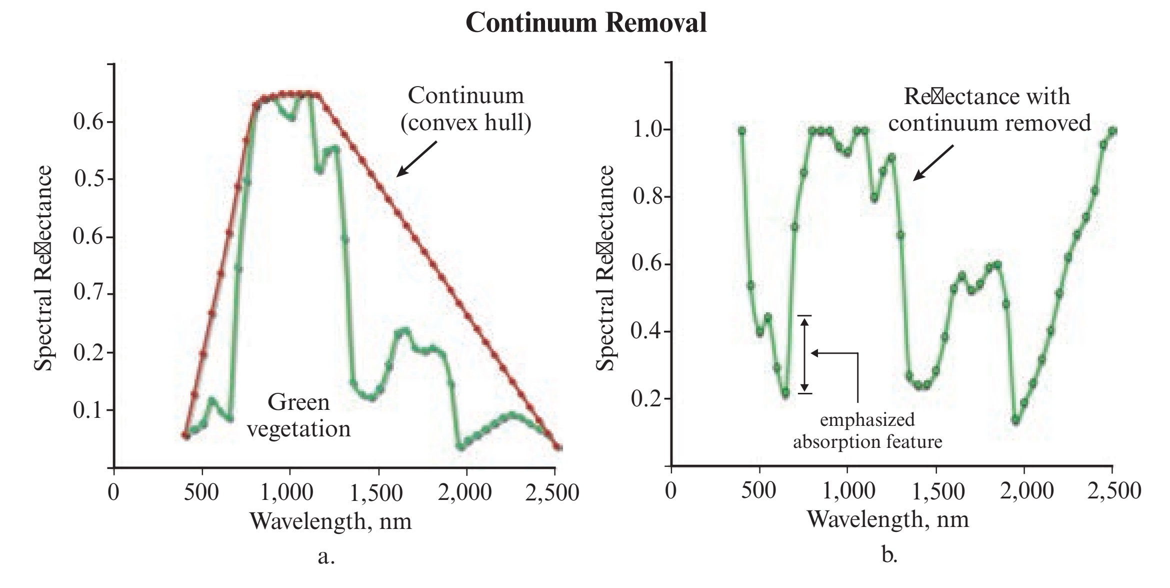


FIGURE 11–20 a) Spectral reflectance curve of healthy green vegetation sampled every 50 nm from 400 to 2,500 nm fitted with a continuum (convex hull) that connects the points of maximum reflection with a straight line. b) Vegetation reflectance after the continuum has been removed using Equation 11.9.

at the ASTER Spectral Library (2014), which contains the:

- Johns Hopkins Spectral Library,
- NASA Jet Propulsion Laboratory Spectral Library, and
- U.S. Geological Survey Spectral Library.

Several digital image processing software vendors provide library spectra as an appendix to their software (e.g., ERDAS, ENVI, PCI Geomatica).

Absorption features are present in the remote sensing– or laboratory-derived reflectance spectra (seen as localized dips) due to the existence of specific minerals, chlorophyll *a* and *b*, water, and/or other materials within the pixel IFOV. The absorption features in the spectra are characterized by their spectral locations (i.e., the bands affected), their depths, and widths.

Considerable attention has been given to developing library matching techniques that allow remote sensing–derived spectra to be compared with spectra that were previously collected in the field or in the laboratory. For example, we may have remote sensing–derived spectra from an agricultural field and desire to compare it with a library containing a significant variety of agricultural spectra. We would compare each pixel's spectrum with each of the agricultural spectra in the library and assign the pixel to the class its spectra most closely resembles.

Comparing the remote sensing–derived spectra with the stored laboratory-derived spectra is no small matter. First, if we are using AVIRIS data, then each pixel has the potential of being represented by 224 distinct spectral band measurements. Each of these distinct measurements would have to be compared with each of the related bands in the library reference spectra. The library may contain hundreds of spectra and the remote sensing scene often consists of millions of pixels. The computation required to conduct true band-to-band comparison between remote sensing–derived spectra and library spectra is daunting. Therefore, various coding techniques have been developed to represent an entire pixel spectrum in a simple yet effective manner so that it can be efficiently compared with library spectra.

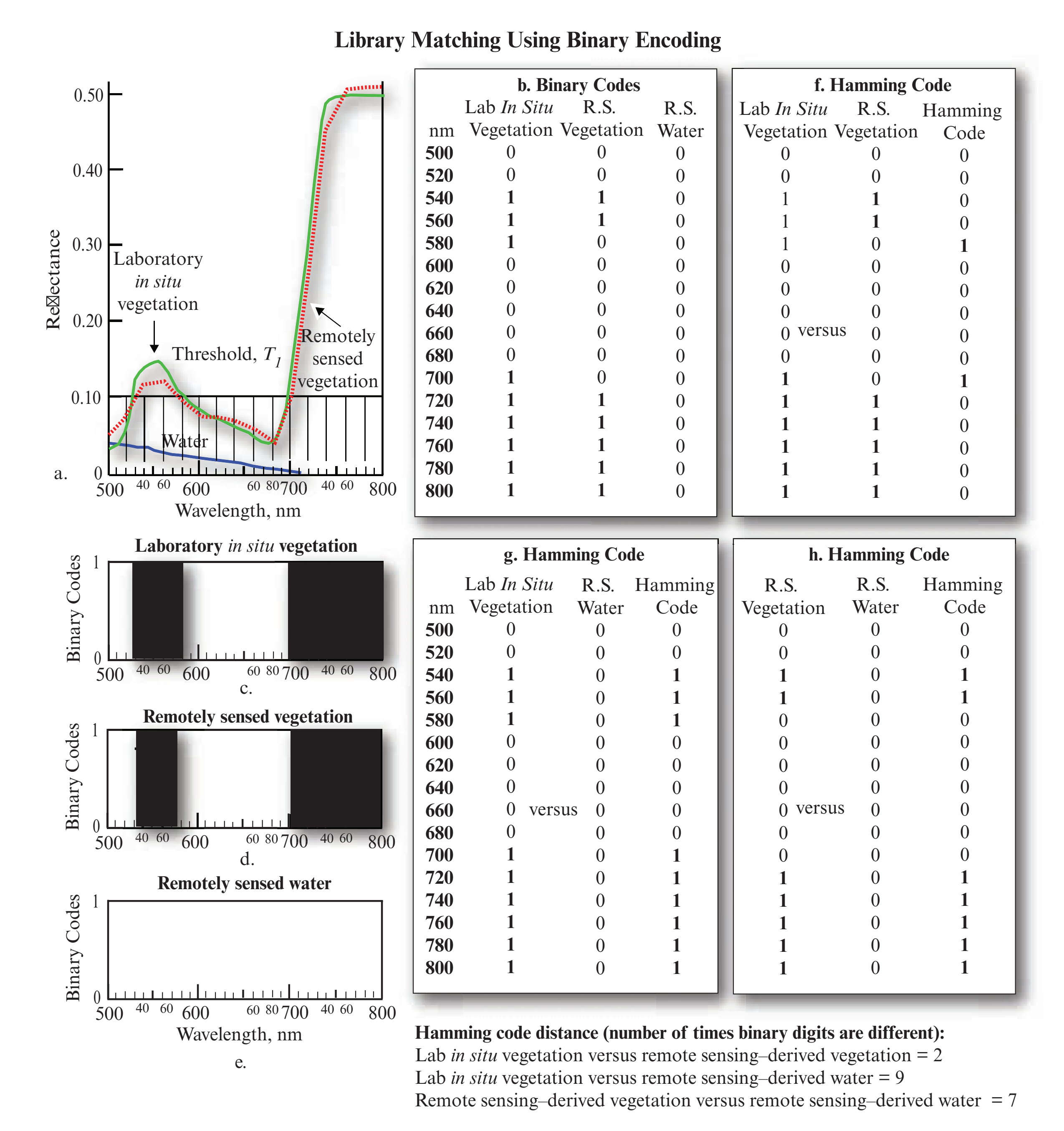
One simple coding technique is *binary spectral encoding* (Jia and Richards, 1993). Binary spectral encoding can be used to transform a hyperspectral reflectance spectrum into simple binary information. The algorithm is:

$$b(k) = 0 \quad \text{if } \rho(k) \le T_1, \tag{11.10}$$

otherwise

$$b(k) = 1$$

where $\rho(k)$ is the reflectance value of a pixel in the kth spectral band, T_1 is a user-selected threshold for creating the binary code, and $b(k)$ is the output binary code symbol for the pixel in the kth spectral band. T_1 may be the average reflectance value of the entire spectrum for a pixel or a user-specified value corresponding to a



b. Binary Codes

nm	Lab *In Situ* Vegetation	R.S. Vegetation	R.S. Water
500	0	0	0
520	0	0	0
540	1	1	0
560	1	1	0
580	1	0	0
600	0	0	0
620	0	0	0
640	0	0	0
660	0	0	0
680	0	0	0
700	1	0	0
720	1	1	0
740	1	1	0
760	1	1	0
780	1	1	0
800	1	1	0

f. Hamming Code

Lab *In Situ* Vegetation		R.S. Vegetation	Hamming Code
0		0	0
0		0	0
1		1	0
1		1	0
1		0	1
0		0	0
0		0	0
0		0	0
0	versus	0	0
0		0	0
1		0	1
1		1	0
1		1	0
1		1	0
1		1	0
1		1	0

g. Hamming Code

nm	Lab *In Situ* Vegetation		R.S. Water	Hamming Code
500	0		0	0
520	0		0	0
540	1		0	1
560	1		0	1
580	1		0	1
600	0		0	0
620	0		0	0
640	0		0	0
660	0	versus	0	0
680	0		0	0
700	1		0	1
720	1		0	1
740	1		0	1
760	1		0	1
780	1		0	1
800	1		0	1

h. Hamming Code

R.S. Vegetation		R.S. Water	Hamming Code
0		0	0
0		0	0
1		0	1
1		0	1
0		0	0
0		0	0
0		0	0
0		0	0
0	versus	0	0
0		0	0
0		0	0
1		0	1
1		0	1
1		0	1
1		0	1
1		0	1

Hamming code distance (number of times binary digits are different):
Lab *in situ* vegetation versus remote sensing–derived vegetation = 2
Lab *in situ* vegetation versus remote sensing–derived water = 9
Remote sensing–derived vegetation versus remote sensing–derived water = 7

FIGURE 11–21 a) *In situ* vegetation spectrum obtained using a spectroradiometer in a controlled laboratory environment (green) and vegetation and water spectra obtained from a remote sensing instrument (red). b) Binary encoding of the *in situ* vegetation, remotely sensed vegetation, and water data. c) Graph of the binary encoding of the *in situ* vegetation data. d) Graph of the binary encoding of the remote sensing–derived vegetation data. e) Graph of the binary encoding of the remote sensing–derived water data. f) Hamming code used in the computation of the Hamming distance between *in situ* vegetation and remote sensing–derived vegetation. g) Hamming code used in the computation of the Hamming distance between *in situ* vegetation and remote sensing–derived water. h) Hamming code used in the computation of the Hamming distance between remote sensing–derived vegetation and water.

unique absorption spectra of particular interest. For example, consider Figure 11-21a, which displays:

- an *in situ* vegetation reflectance spectrum measured in a laboratory using a spectroradiometer, and
- vegetation and water spectra obtained using an airborne hyperspectral sensor.

A threshold of 10% reflectance was chosen because it captured nicely the chlorophyll *a* and *b* absorption bands in the blue and red portions of the spectrum. With the threshold set at 10%, it was a straightforward

task to binary encode each of the three spectra and then compare their binary codewords to see how similar or different they were. For example, the *in situ* vegetation spectra exceeded the threshold in the region from 540 to 580 nm and again in the region from 700 to 800 nm. Therefore, each of these wavelength regions (bands) received a value of 1 in their codewords in Figure 11-21b. The remote sensing–derived vegetation spectra had very similar characteristics, exceeding the threshold from 540 to 560 nm and from 720 to 800 nm. Once again, all these bands were coded to 1. The remote sensing–derived water spectra never exceeded the threshold and therefore received zeros in all bands.

Thus, each of the original spectra was turned into a 16-digit codeword, c_1, c_2, and c_3. But how do we use these three binary codewords (each of length $Z = 16$) to determine if the binary spectra of one class is identical, similar to, or completely different from the spectra of another class? Two spectra that have been binary encoded can be compared by computing the Hamming distance between their binary codewords using:

$$\text{Dist}_{Ham}(c_i, c_j) = \sum_{k=1}^{N = \text{bands}} [c_i(k) \Theta c_j(k)] \quad (11.11)$$

where c_i and c_j are two spectral codewords of length L and the Θ symbol represents exclusive OR Boolean logic. The algorithm compares each bit of the two codewords under examination and outputs a 0 where they agree and a 1 where they are different. The result is a Hamming distance codeword of length Z.

Thus, the Hamming distance is computed by summing the number of times the binary digits are different. For example, consider the computation of the Hamming distance between the laboratory *in situ* vegetation codeword and the remote sensing–derived vegetation codeword. The two codewords differ at only two locations in the 16-bit lengths and therefore have a Hamming distance of 2. Conversely, when the laboratory *in situ* vegetation codeword is compared with the water codeword, it differs at nine locations, resulting in a Hamming distance of 9.

The more similar one spectra is to another, the lower its Hamming distance. Thus, one can a) identify the appropriate threshold for a material or land-cover class of interest in light of existing absorption bands, b) compute a binary codeword at each pixel, and c) compare this binary codeword at each pixel with *m* library spectra that have been binary coded in a similar manner. Each pixel could be assigned to the class that results in the smallest Hamming distance.

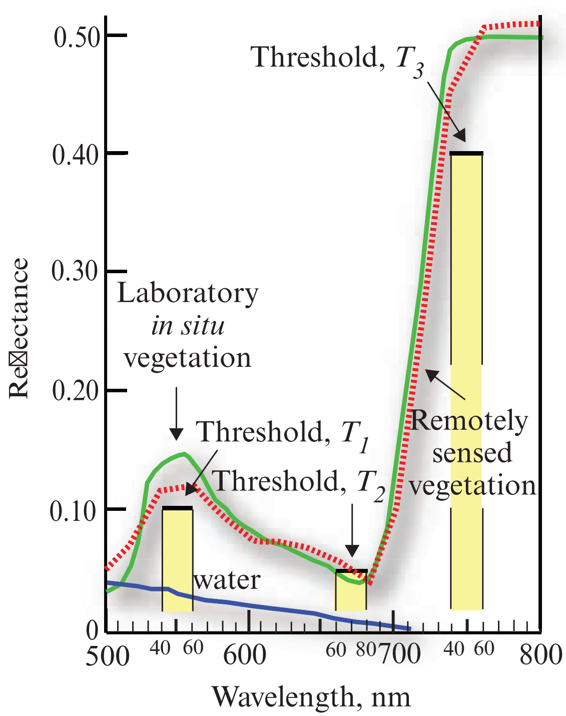


FIGURE 11–22 The location of three thresholds that could be used to perform binary encoding of the three spectra from 540 to 560 nm, 660 to 680 nm, and 740 to 760 nm set at 10%, 5%, and 40% reflectance, respectively.

Note that more than one threshold can be used a) within a chosen absorption band, and b) at numerous disjoint locations along the spectra. For example, Figure 11-22 depicts the location of three unique thresholds that could be used to perform binary encoding using just the information at T_1 (540 to 560 nm), T_2 (660 to 680 nm), and T_3 (740 to 760 nm) set at 10%, 5%, and 40% scaled reflectance, respectively.

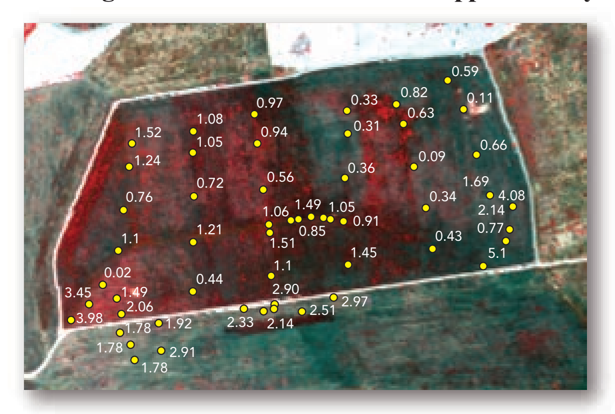
Machine Learning Analysis of Hyperspectral Data

Scientists have also used various types of machine-learning algorithms to analyze hyperspectral data such as decision trees and support vector machines (SVM).

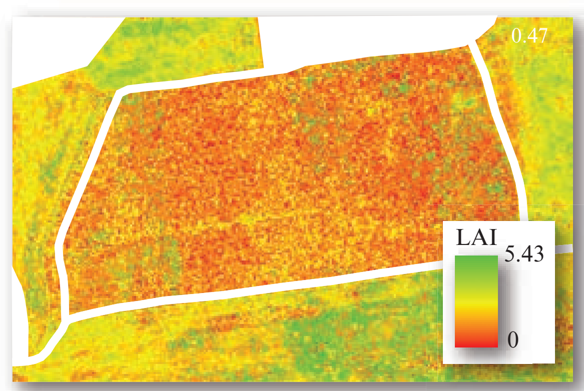
Decision Tree Analysis of Hyperspectral Data

As described in Chapter 10, decision trees can be used to map the spatial distribution of nominal-scale land cover (e.g., urban, wetland, forest) and regression trees can be used to extract quantitative information about continuous variables (e.g., LAI, biomass, percent canopy cover) (e.g., Im and Jensen, 2008).

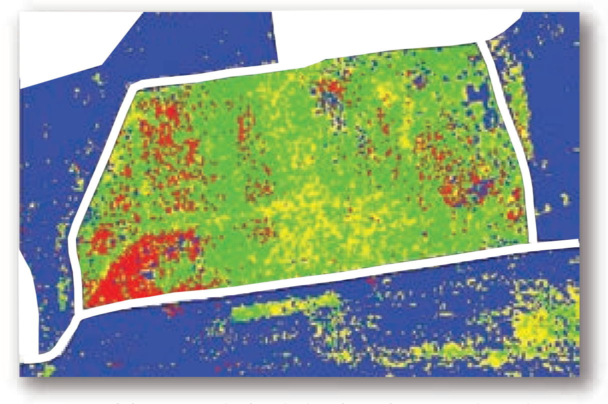
Mapping LAI and Land Cover of a Capped Waste Site near Monticello, UT, using DecisionTree– and Regression Tree–Derived Rules Applied to HyMap Hyperspectral Data acquired on June 2, 2008



a. HyMap hyperspectral imagery (RGB = bands 24, 17, 11) with 2.3 × 2.3 m spatial resolution overlaid with the location of 53 *in situ* LAI measurements.



b. Predicted LAI distribution based on regression tree–rules derived from vegetation indices and mixture-tuned matched filtering (MTMF) derived metrics.



Big Sagebrush (*Artemisia tridentata*)
Rubber Rabbitbrush (*Ericameria nauseosa*)
Western Wheatgrass (*Pascopyrum smithii*)
Litter (primarily dead grass)

c. Decision tree–derived classification map based on analysis of vegetation indices and mixture-tuned-matched filtering (MTMF) derived metrics.

FIGURE 11–23 a) Color-composite of three of the 126 HyMap hyperspectral bands overlaid with the location of 53 *in situ* sampling locations where LAI and land-cover information were collected. b) LAI spatial distribution based on regression-tree machine learning applied to vegetation indices and MTMF-derived metrics. c) Land-cover classification based on decision-tree analysis applied to vegetation indices and MTMF-derived metrics (based on Im, J., Jensen, J. R., Jensen, R. R., Gladden, J., Waugh, J. and M. Serrato, 2012, "Vegetation Cover Analysis of Hazardous Waste Sites in Utah and Arizona using Hyperspectral Remote Sensing," *Remote Sensing*, 2012(4):327–353).

Example 1 Im et al. (2012) used *decision tree*–derived rules to map land cover and *regression tree*–derived rules to map the LAI on an engineered waste site in Monticello, UT. HyMap airborne hyperspectral data were collected in 126 bands from 440 – 2,500 nm at 2.3 × 2.3 m spatial resolution on June 2–3, 2008 (Figure 11-23a). The HyMap imagery were atmospherically corrected to scaled percent reflectance and polished using EFFORT. The polished scaled reflectance data were geometrically rectified to a Universal Transverse Mercator (UTM) projection using 20 Ground Control Points (GCPs) collected from 2006 National Agricultural Imagery Program (NAIP) digital orthophotos (1 × 1 m spatial resolution) resulting in a root mean square error (RMSE) of <1 pixel.

LAI and land-cover information were collected *in situ* at the time of the HyMap overflight at the 53 locations shown in Figure 11-23a. The dominant species included Big Sagebrush (*Artemisia tridentata*), Rubber Rabbitbrush (*Ericameria nauseosa*), Western Wheatgrass (*Pascopyrum smithii*) and dead litter plant material (Table 11-4).

Several information extraction techniques were investigated in this study but the machine-learning decision-tree classifiers yielded in the best results. A machine learning regression tree (Cubist) was used to analyze hyperspectral vegetation indices and mixture-turned matched filtering (MTMF) derived metrics to generate a rule-based model of LAI. The usefulness of Cubist for creating robust regression trees is well documented in the remote sensing literature (e.g., Huang and Townshend, 2003; Moisen et al., 2006; Im et al., 2009).

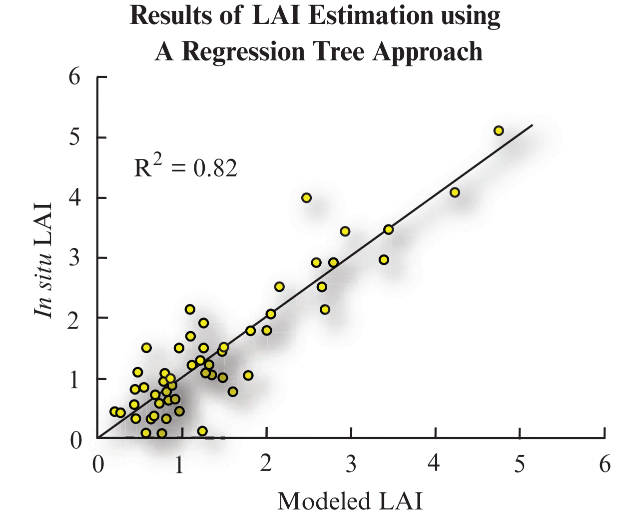


FIGURE 11–24 A scatterplot of 53 *in situ* reference LAI measurements versus LAI measurements estimated using a regression tree approach applied to HyMap imagery for a waste site near Monticello, UT (based on Im et al., 2012, as cited in Figure 11-23).

TABLE 11–4 Field data characteristics for vegetation mapping and leaf-area-index (LAI) estimation on the Monticello, UT, waste site in June, 2008 (based on Im et al., 2012, as cited in Figure 11-23).

Classes	Number of Samples
Big Sagebrush (*Artemisia tridentata*)	8
Rubber Rabbitbrush (*Ericameria nauseosa*)	12
Western Wheatgrass (*Pascopyrum smithii*)	16
Litter (primarily dead grass)	17
LAI ranged from 0.09–5.43	53

LAI estimation using the regression tree approach for the Monticello, UT, study area resulted in an R^2 of 0.82. A map of the spatial distribution of LAI is shown in Figure 11-23b. A scatterplot of the *in situ* LAI and the modeled LAI is shown in Figure 11-24.

A vegetation species classification map created using the decision-tree classifier is shown in Figure 11-23c. The decision tree used various vegetation indices plus mixture-tuned matched-filtering (MTMF) derived variables to generate the rules. The overall classification accuracy was 86.8% with a Kappa Coefficient of Agreement of 0.82. The contingency table is provided in Im et al. (2012). The study concluded that monitoring vegetation cover on waste sites using hyperspectral remote sensing data and decision-tree machine learning is practical although higher spatial resolution imagery (<1 × 1 m) is preferable.

Example 2 The previous study focused on the identification of LAI and land cover. It is also possible to analyze hyperspectral remote sensor data to obtain data on other biophysical and biochemical measurements. For example, Im et al. (2009) used high spatial resolution (1 × 1 m) hyperspectral data of experimental short rotation woody crop plots on the Savannah River Site near Aiken, SC, to extract detailed biophysical and biochemical information (Figure 11-25a).

The objectives of this study were to (1) identify the relationship between *in situ* and hyperspectral remote sensing measurement in terms of tree species and the different levels of water and nutrient availability, and (2) estimate biophysical and biochemical characteristics (i.e. LAI, biomass, leaf nutrients concentrations) from the hyperspectral imagery using three different digital image processing techniques.

The leaf nutrient variables (treatments) included nitrogen (N), phosphorous (P), potassium (K), calcium (Ca), and magnesium (Mg). The randomized block design of the treatments is shown in Figure 11-25b. Treatments were composed of (1) irrigation only, (2) irrigation and fertilization, (3) control (no irrigation and fertilization), and (4) fertilization only. Different fertilization levels were also applied to experimental plots in blocks 1 and 5, but they were not included in this study due to limited *in situ* measurements. A total of 60 plots in blocks 2–4 were used, which included 20 treatment plots (5 genotypes grown with 2 fertilizer and 2 irrigation treatments) with each replicated three times (Figure 11-25b). Detailed information about the experimental design is found in Coleman et al. (2004).

An Airborne Imaging Spectrometer for Applications (AISA) Eagle sensor system collected hyperspectral imagery over the study area on September 15, 2006 (Figure 11-25a). The imagery consisted of 63 channels (from 400 to 980 nm) with a spectral resolution of approximately 9-nm, a radiometric resolution of 12-bits, and a spatial resolution of 1 × 1 m. The imagery was collected at an altitude of 1,630 m above ground level (AGL) during cloud-free conditions at 11:20 a.m. (EDT) local time.

The hyperspectral imagery were pre-processed to scaled percent reflectance using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercube (FLAASH) algorithm and rectified to a Universal Transverse Mercator (UTM) coordinate system. The geometric rectification was performed using numerous GPS-derived coordinates located in the study site (i.e., each corner of each block) resulting in an estimated Root Mean Square Error (RMSE) of 0.48 pixels.



Treatment	Fertilization	Irrigation	# plots
1	No	Yes	15
2	120	Yes	15
3	No	No	15
4	120	No	15
5	30	Yes	3
6	60	Yes	5
7	90	Yes	3
8	150	Yes	3

Treatment	Fertilization	Irrigation	# plots
9	180	Yes	3
10	210	Yes	3
11	20	No	2
12	40	No	2
13	60	No	5
14	80	No	2
15	100	No	2
16	140	No	2

FIGURE 11–25 a) Color-composite of AISA hyperspectral data at 1 × 1 m spatial resolution of short rotation woody crop experimental plots. b) Characteristics of the randomized block design associated with vegetation type and fertilization and irrigation treatments. c) Characteristics of the 60 plots used in the study (based on Im, J., Jensen, J. R., Coleman, M., and E. Nelson, 2009, "Hyperspectral Remote Sensing Analysis of Short Rotation Woody Crops Grown with Controlled Nutrient and Irrigation Treatments," *Geocarto International*, 24(4):293–312).

TABLE 11–5 Results of predicting biophysical and biochemical characteristics of short rotation woody crops using AISA Eagle hyperspectral imagery (based on Im, J., Jensen, J. R., Coleman, M. and E. Nelson, 2009, "Hyperspectral Remote Sensing Analysis of Short Rotation Woody Crops Grown with Controlled Nutrient and Irrigation. Treatments," *Geocarto International*, 24(4):293–312).

Target Variable	Calibration R^2
NDVI (using bands centered on 657.0 and 856.8 nm)	
LAI	0.738
Biomass	0.569
N	0.016
P	0.019
K	0.002
Ca	0.225
Mg	0.457
Partial Least Squares Regression (using 63 reflectance variables	
LAI	0.854
Biomass	0.633
N	0.133
P	0.177
K	0.222
Ca	0.484
Mg	0.506
Machine-learning Regression Tree (using 63 reflectance variables	
LAI	0.956
Biomass	0.855
N	0.793
P	0.229
K	0.557
Ca	0.626
Mg	0.635

The three digital image processing techniques used were (1) simple regression using normalized difference vegetation index (NDVI), (2) partial least squares regression, and (3) machine-learning regression tree. Results of the hyperspectral analysis using the three techniques are summarized in Table 11-5 (Im et al., 2009).

This study demonstrated the capability of hyperspectral remote sensing to predict selected forest biophysical and biochemical characteristics in a very complex environment containing five tree genotypes subjected to different levels of irrigation and/or fertilization. The simple linear regression based on NDVI resulted in good performance in predicting the biophysical characteristics such as LAI and stem biomass for certain genotypes. But, it failed to predict the biochemical characteristics (i.e., leaf nutrients) covering all tree species and treatments (Table 11-5). The partial least squares regression and the regression-tree analysis yielded much better estimation of the leaf nutrients concentrations (Table 11-5). In particular, the regression-tree approach appears to be a robust method for predicting biophysical and biochemical properties from environmental complex forest environments when provided with sufficient high-quality training data.

Support Vector Machine (SVM) Analysis of Hyperspectral Data

The SVM classification algorithm described in Chapter 10 is an effective distribution free classifier that has been widely used in the recent years for solving hyperspectral classification problems (Camps-Valls and Bruzzone, 2005; Dalponte et al., 2009). The main reason for the choice of this classifier is associated with its properties, including: a) high generalization ability and high classification accuracies (with respect to other classifiers); b) effectiveness in addressing *ill-posed* problems (which are quite common with hyperspectral data mainly because of the limited amount of training samples relative to the number of features (bands) (Plaza et al., 2009); and c) limited effort required for architecture design and training phase when compared to other machine-learning algorithms (such as multilayer perceptron neural networks). Dalponte et al. (2009) compared SVM, maximum-likelihood, and discriminant analysis classification techniques for spectral resolutions ranging from 4.6 – 36.8 nm and found that SVM consistently yielded the highest classification accuracy, especially when using the highest spectral resolution hyperspectral data (4.6 nm). Plaza et al. (2009) used a suite of SVM to analyze hyperspectral data using a variety of techniques that incorporate both spatial and hyperspectral data.



Selected Indices Useful for Hyperspectral Data Analysis

Traditional vegetation indices (discussed in Chapter 8) developed for use with broadband multispectral data can be used with hyperspectral data. In fact, Blonski et al. (2002) developed an algorithm that will synthesize multispectral bands (e.g., Landsat ETM^+ bands) from hyperspectral data (e.g., AVIRIS), if desired. Several studies identify a variety of indices developed especially for use with hyperspectral data for a variety of applications (e.g., Im et al., 2009; Im et al., 2012; Brantley et al., 2011). Thenkabail et al. (2000; 2004; 2011), Haboudane et al. (2002), Stagakis et al. (2010) and Behmann et al. (2014) review several narrow-band indices for remote sensing of agriculture and other types of vegetation. Ustin et al. (2009) review indices for retrieving plant pigment information using high-resolution spectroscopy. Matthews (2011) provides an overview of indices used for remote sensing inland and near-coastal waters. Alonzo et al. (2014) describe how hyperspectral data are used with LiDAR data to perform urban tree mapping. Jay et al. (2014) used hyperspectral data to map water quality and depth. Selected indices that can be applied to hyperspectral data are provided below.

Reduced Simple Ratio

Brown et al. (2000) introduced the Reduced Simple Ratio (RSR):

$$RSR = \frac{\rho_{nir}}{\rho_{red}}\left(1 - \frac{\rho_{swir} - \rho_{swirmin}}{\rho_{swirmax} - \rho_{swirmin}}\right) \quad (11.12)$$

where $\rho_{swirmin}$ and $\rho_{swirmax}$ are the scene minimum and maximum short wavelength infrared (*swir*) reflectance. Chen et al. (2002) and Gray and Song (2012) used the RSR for mapping leaf-area-index.

Normalized Difference Vegetation Index—NDVI

As discussed in Chapter 8, the normalized difference vegetation index is based on the equation:

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}. \quad (11.13)$$

Generally, reflectances from a red channel centered around 660 nm and a near-infrared channel centered on 860 nm are used to calculate the NDVI. The near-infrared band empirically corresponds to the long-wavelength shoulder of the chlorophyll red-edge, and the red band is associated with the maximum chlorophyll absorption. The following is one possible narrow-band implementation of the standard NDVI:

$$NDVI_{narrow} = \frac{\rho_{860} - \rho_{660}}{\rho_{860} + \rho_{660}}. \quad (11.14)$$

A narrow-band NDVI image derived from an analysis of AVIRIS bands 29 (red) and 51 (near-infrared) is shown in Figure 11-26a. As expected, the lower biomass areas in the NDVI image correspond generally to stressed areas in the thematic map.

Hyperspectral Enhanced Vegetation Index—EVI

Clark et al. (2011) used the following narrow-band hyperspectral Enhanced Vegetation Index:

$$EVI_{hyper} = \frac{\rho_{798} - \rho_{679}}{[1 + \rho_{798} + (6 \times \rho_{679}) - (7.5 \times \rho_{482})]} \quad (11.15)$$

applied to HYDICE imagery to extract photosynthesizing vegetation structure, senescence, and health information. Behmann et al. (2014) used hyperspectral bands in an EVI to assess vegetation stress.

Yellowness Index—YI

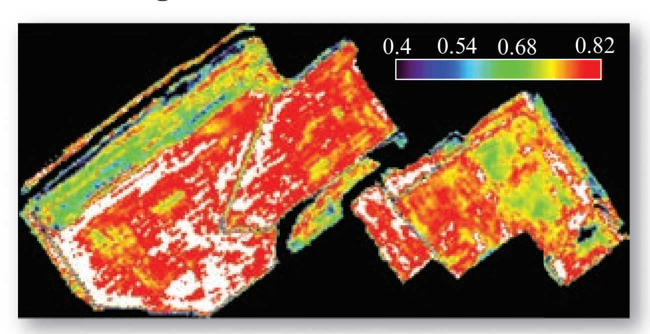
The yellowness index (YI) constitutes a descriptor of leaf chlorosis exhibited by stressed vegetation. It measures the change in shape of the reflectance spectra in the interval between the 0.55-μm reflectance maximum and the 0.65-μm minimum. The YI is a three-point gross measure of green–red spectral shape and is thus computed using only wavelengths in the visible spectrum. The justification is that the specified visible region tends to be relatively insensitive to changing leaf water content and structure (Philpot, 1991; Adams et al., 1999):

$$YI = \frac{\rho(\lambda_{-1}) - 2\rho(\lambda_0) + \rho(\lambda_{+1})}{\Delta\lambda^2} = \frac{d^2\rho}{d\lambda^2} \quad (11.16)$$

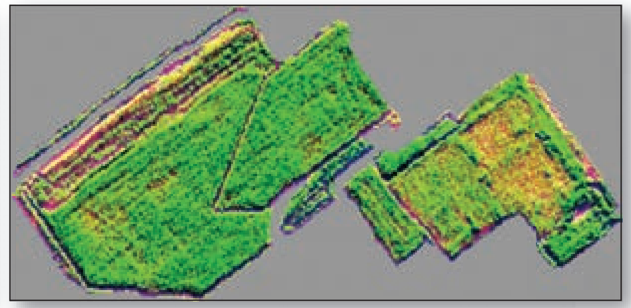
where $\rho(\lambda_0)$ is the band center reflectance; $\rho(\lambda_1)$ and $\rho(\lambda_{+1})$ denote the reflectance of the lower and higher wavebands, respectively; and $\Delta\lambda$ is the spectral distance (measured in μm) between wavebands (i.e., $\Delta\lambda = \lambda_0 - \lambda_{-1} = \lambda_{+1} - \lambda_0$). The goal is to select wavelengths such that the band separation ($\Delta\lambda$) is as large as possible while constraining all three channels in the spectral range between approximately 0.55 μm and 0.68 μm. The YI can be computed as 0.1 multiplied by the negative of the finite approximation of the second derivative as a means of downscaling the range of output values; thus, the resultant relationship would be that as yellowness increases, so do positive YI values (Adams et al., 1999). The YI is in units of relative reflectance μm^{-2} (RRU μm^{-2}). The YI magnitude is sensitive to the λ_c and $\Delta\lambda$ values. Using 1999 AVIRIS data, each value in the YI equation can be the average of three adjacent bands; λ_0 centered at 0.61608 μm (band 26); λ_{-1} centered at 0.56696 μm (band 21); and λ_{+1} centered at 0.66518 μm (band 31) based on Adams et al. (1999).

Physiological Reflectance Index—PRI

The physiological reflectance index (PRI) is a narrow-band index that has been correlated with the epoxidation state of the xanthophyll pigments and with photosynthetic efficiency with respect to control (unstressed) and nitrogen-stressed canopies (Gamon et al., 1992). However, the PRI is generally not well correlated with water-stressed canopies experiencing midday wilting. The PRI employs the reflectance at 531 nm and a reference channel in order to minimize the effects of diurnal Sun angle changes. The PRI can track diurnal changes in photosynthetic efficiency. In addition, the PRI may be useful in situations where *in situ* spectroradiometer data are acquired at a different time or light regime geometry than the remote sensor data. The general PRI formula is:



a. NDVI image derived from AVIRIS bands 29 (0.64554 μm) and 51 (0.82593 μm).



b. Second-order derivative image.

FIGURE 11–26 a) Normalized difference vegetation index (NDVI) image derived from AVIRIS bands 29 (0.64554 μm) and 51 (0.82593 μm), red and near-infrared, respectively. White and red pixels indicate the greatest amount of biomass. Blue, green, and orange pixels contain less biomass. b) A second-order derivative image extracted from the original AVIRIS scaled surface reflectance data (RGB = bands 42, 30, 18). Original AVIRIS imagery courtesy of NASA.

$$PRI = \frac{\rho_{ref} - \rho_{531}}{\rho_{ref} + \rho_{531}} \quad (11.17)$$

where ρ_{ref} is a reference wavelength and ρ_{531} is the reflectance at 531 nm. The best PRI noted in Gamon et al. (1992) was:

$$PRI = \frac{\rho_{550} - \rho_{531}}{\rho_{550} + \rho_{531}}. \quad (11.18)$$

A reference wavelength of 550 nm seems to be appropriate at the canopy level. For leaf-scale spectra, a reference wavelength of 570 nm is likely better for a xanthophyll signal. Note that a single PRI is not likely to be applicable across all spatial and temporal scales as well as various canopy types and diurnally dynamic canopy structures. AVIRIS implementation for the 1999 data might include bands 19 (centered at 547.32 nm) and 17 (centered at 527.67 nm), respectively.

In contrast with the PRI, the NDVI does not accurately indicate real-time photosynthetic fluxes. While NDVI is often sensitive at low leaf-area-index (LAI) values, it often saturates at high LAI values. The PRI may be able to indicate short-term changes in photosynthetic efficiency, especially in canopies with high LAI values where the NDVI is least effective (Gamon et al., 1992).

Normalized Difference Water Index —NDWI

The normalized difference water index (NDWI) may be used to remotely determine vegetation liquid water content. Two near-infrared channels are used in the computation of the NDWI; one is centered at approximately 860 nm and the other at 1,240 nm (Gao, 1996; Gray and Song, 2012):

$$NDWI = \frac{\rho_{860} - \rho_{1240}}{\rho_{860} + \rho_{1240}}. \quad (11.19)$$

AVIRIS bands that might be used include bands 55 (864.12 nm) and 94 (1237.94 nm). Clark et al. (2011) used HYDICE wavelengths of 862 and 1,239 nm.

Linear Red-edge Position—REP

The red-edge position (REP) is defined as the point of maximum slope on a vegetation reflectance spectrum between the red and near-IR wavelengths. The REP is useful because it is strongly correlated with foliar chlorophyll content and can be a sensitive indicator of vegetation stress. Although the AVIRIS sensor has a nominal band separation of 10 nm, subtle REP shifts may not be discernible because the spectra are still sampled coarsely relative to data from laboratory spectrometers (Dawson and Curran, 1998).

A linear method proposed by Clevers (1994) can be implemented that makes use of four narrow bands and is computed as (Im et al., 2012):

$$REP = 700 + 40\left[\frac{\rho_{red\ edge} - \rho_{700}}{\rho_{740} - \rho_{700}}\right] \quad (11.20)$$

where

$$\rho_{red\ edge} = \frac{\rho_{670} + \rho_{780}}{2}. \qquad (11.21)$$

A derivative-based REP algorithm exists that is of low computational complexity and is appropriate for some canopy-scale studies (Dawson and Curran, 1998).

Red-Edge Vegetation Stress Index (RVSI)

Clark et al. (2011) used the following narrow-band hyperspectral Red-edge Vegetation Stress Index (RVSI)

$$RVSI = \left[\left(\frac{\rho_{719} + \rho_{752}}{2}\right)\right] - \rho_{730} \qquad (11.22)$$

applied to HYDICE imagery to extract photosynthesizing vegetation stress information.

Crop Chlorophyll Content Prediction

Haboudane et al. (2002) developed a narrow-band vegetation index that integrates the advantages of indices that minimize soil background effects and indices that are sensitive to chlorophyll concentration. The Transformed Chlorophyll Absorption in Reflectance Index (TCARI) (Daughtry et al., 2000) is:

$$TCARI = 3\left[(\rho_{700} - \rho_{670}) - 0.2(\rho_{700} - \rho_{550})\left(\frac{\rho_{700}}{\rho_{670}}\right)\right]. \qquad (11.23)$$

The Optimized Soil-Adjusted Vegetation Index (Rondeaux et al., 1996) belongs to the Soil-Adjusted Vegetation Index (SAVI) family (Huete, 1988) and is:

$$OSAVI = \frac{(1 + 0.16)(\rho_{800} - \rho_{670})}{(\rho_{800} + \rho_{670} + 0.16)}. \qquad (11.24)$$

The ratio

$$\frac{TCARI}{OSAVI} \qquad (11.25)$$

is sensitive to chlorophyll content variations and resistant to variations of leaf-area-index (LAI) and solar zenith angle. Evaluation with ground truth resulted in an $r^2 = 0.81$ between estimated and field-measured chlorophyll content data. Haboudane et al. (2002) suggest that the ratio index is insensitive to LAI variations for LAI values ranging from 0.5 to 8 and that it might be used operationally in precision agriculture because it allows an accurate estimation of crop photosynthetic pigments without *a priori* knowledge of crop canopy architecture.

Modified Chlorophyll Absorption Ratio Index (MCARI1)

Haboudane et al. (2002) proposed the modified chlorophyll absorption ratio index (MCARI1):

$$MCARI1 = 1.2[2.5(\rho_{800} + \rho_{670}) - 1.3(\rho_{800} - \rho_{550})] \qquad (11.26)$$

which is more sensitive to variations in green LAI.

Chlorophyll Index

Brantley et al. (2011) used a chlorophyll index (CI) based on the work of Gitelson et al. (2005) that makes use of red-edge wavelengths:

$$CI = \frac{\rho_{750}}{(\rho_{700} + \rho_{710}) - 1}. \qquad (11.27)$$

It is a good predictor of chlorophyll content for vegetation with high LAI.

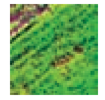
Medium Resolution Imaging Spectrometer (MERIS) Terrestrial Chlorophyll Index (MTCI)

Dash et al. (2010) used a European Space Agency Envisat product, the Medium Resolution Imaging Spectrometer (MERIS) Terrestrial Chlorophyll Index (MTCI), to study the spatio-temporal variation of vegetation phenology over India:

$$MTCI = \frac{MERIS_{band10} - MERIS_{band9}}{MERIS_{band9} - MERIS_{band8}} \qquad (11.28)$$

$$MTCI = \frac{\rho_{753.75} - \rho_{708.75}}{\rho_{708.75} - \rho_{681.25}}.$$

This index is highly correlated with chlorophyll concentration and leaf-area-index and has limited sensitivity to atmospheric effects, soil background, and view angle.



Derivative Spectroscopy

Differentiation of a curve or its mathematical function estimates the slope over the entire interval. Deriving the slope of spectroradiometer-derived curves is called *derivative spectroscopy*. Derivative spectroscopy methods were originally developed in analytical chemistry to eliminate background signals and resolve overlapping spectral features (Demetriades-Shah et al., 1990).

The concept has also been applied to differentiating remote sensing–derived spectra. Differentiation does not yield more information than exists in the original spectral channels. It can be used, however, to emphasize desired information while suppressing or eliminating information not of interest. For example, background absorption or reflectance signals caused by stray light can be eliminated (Talsky, 1994). Spectral features with sharp structures may be enhanced compared with broader-structured features.

Because of the inherent differences between lab- and remote sensor–based data, not all lab-based spectroscopic procedures translate well to remotely sensed inquiries. Lab analysis is characterized by controlled illumination sources and viewing geometries, as well as by the typical assumption of homogenous target samples and the use of a known standard. Conversely, imaging spectroscopic remote sensing entails a natural illumination source, mixed pixels, varying topography, generally coarser spectral resolution than that of laboratory spectrophotometers, and the general lack of useful reference standards (Tsai and Philpot, 1998). Given these differences, the application of laboratory spectroscopic techniques to remote sensor data must be carefully considered. Derivative techniques have been applied to remote sensor imagery, but such applications have been rather limited to date (e.g., Demetriades-Shah et al., 1990; Philpot, 1991; Li et al., 1993; Penuelas et al., 1993; Tsai and Philpot, 1998; Thenkabail et al., 2004; Mitchell et al., 2012). Nevertheless, there are advantages to using derivative-based analytical techniques with remotely sensed data. Derivative spectra are sometimes more informative than zero-order reflectance spectra when attempting to glean certain information or relationships from the data. For example, Malthus and Madeira (1993) found the first-order derivative spectra in the visible wavelengths to be more highly correlated with percent leaf surface area infected by the fungus *Botrytis fabae* than the original zero-order reflectance data.

First-, second-, and third-order derivative spectra may be computed on a pixel-by-pixel basis. The differentiation can be performed using three-point Lagrangian formulas (Hildebrand, 1956):

$$a'_{-1} = \frac{1}{2h}(-3a_{-1} + 4a_0 - a_1) + \frac{h^2}{3}a(\xi) \qquad (11.29)$$

$$a'_0 = \frac{1}{2h}(-a_{-1} + a_1) - \frac{h^2}{6}a(\xi) \qquad (11.30)$$

$$a'_1 = \frac{1}{2h}(a_{-1} - 4a_0 + 3a_1) + \frac{h}{3}a(\xi) \qquad (11.31)$$

where the subscripts 0, –1, and 1 denote the first derivative at the center point, and points to the left and right of the center point, respectively. h and ξ are distance and error terms, respectively. Higher-order derivative spectra are calculated in an analogous manner.

In remote sensing investigations, higher-order derivatives (second-order and higher) are relatively insensitive to illumination intensity variations, due to cloud cover, Sun angle variance, or topographic effects. In addition, derivatives are usually insensitive to changes in spectral solar flux and skylight (Tsai and Philpot, 1998). Derivative techniques can also be used to address interference from soil background reflectance in vegetation remote sensing studies (i.e., separating the vegetative signal from the background noise). Second-order derivative spectra, which are insensitive to soil reflectance, specifically mitigate this problem, while first-order derivative spectra do not (Demetriades-Shah et al., 1990; Li et al., 1993).

Narrow-Band Derivative-Based Vegetation Indices

A variety of narrow-band derivative-based vegetation indices may be computed. For instance, several derivative vegetation indices that measure the amplitude of the chlorophyll red-edge using hyperspectral data in the 626-nm to 795-nm spectral range can be implemented (Elvidge and Chen, 1995):

$$\text{1DL_DGVI} = \sum_{\lambda_1}^{\lambda_n} \left|\rho'(\lambda_i) - \rho'(\lambda_1)\right| \Delta\lambda_i \qquad (11.32)$$

$$\text{1DZ_DGVI} = \sum_{\lambda_1}^{\lambda_n} \left|\rho'(\lambda_i)\right| \Delta\lambda_i \qquad (11.33)$$

$$\text{2DZ_DGVI} = \sum_{\lambda_1}^{\lambda_n} \left|\rho''(\lambda_i)\right| \Delta\lambda_i \qquad (11.34)$$

$$\text{3DZ_DGVI} = \sum_{\lambda_1}^{\lambda_n} \left|\rho'''(\lambda_i)\right| \Delta\lambda_i \qquad (11.35)$$

where i is the band number; λ_i is the center wavelength for the ith band; λ_1 = 626 nm; λ_2 = 795 nm; ρ' is the first derivative reflectance; and ρ'' is the second derivative reflectance. Higher-order derivative-based indices are also possible. For instance, ρ''' is the third-order derivative reflectance. Thenkabail et al. (2004) used this index to determine the best hyperspectral wavebands for studying vegetation and agricultural crops.

The 16-bit radiometric resolution of AVIRIS data facilitates the creation of quality higher-order derivative spectra. For example, a second-order derivative image of the MWMF on the Savannah River Site is displayed in Figure 11-26b. Approximately the same areas of healthy vegetation and potential stressed vegetation are apparent in the derivative image, but they are more visually striking.

Mitchell et al. (2012) analyzed HyMap hyperspectral data using derivative analysis to map semiarid shrubland Nitrogen (N) content. They were able to quantify sagebrush canopy N concentrations using partial least squares regression with an R^2 of 0.72.

Red-Edge Position Based on Derivative Ratio

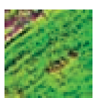
Smith et al. (2004) used a red-edge positioning approach where reflectance data were first smoothed using a weighted mean moving average function:

$$\rho_{red\text{-}edge} = \frac{(\rho_{726_{smooth}} + \rho_{724_{smooth}})/2}{(\rho_{703_{smooth}} + \rho_{701_{smooth}})/2} \qquad (11.36)$$

where

$$\rho_{\lambda_{i,smooth}} = \frac{0.25 \cdot \rho_{\lambda_{i-2}} + 0.5 \cdot \rho_{\lambda_{i-1}} + \rho_{\lambda_i} + 0.5 \cdot \rho_{\lambda_{i+1}} + 0.25 \cdot \rho_{\lambda_{i+2}}}{0.25 + 0.5 + 1 + 0.5 + 0.25}$$

The function used a 5-nm sample range, which consists of five values along the reflectance continuum. The relative weights of 0.25, 0.5, 1, 0.5, and 0.25 are applied to the values of the reflectance continuum to compute the average value. The derivative is then calculated by dividing the difference between two average reflectance values with a 2-nm interval. Additional methods used to determine the red-edge position are summarized in Baranoski and Rokne (2005).



References

Adams, M. L., Philpot, W. D., and W. A. Norvell, 1999, "Yellowness Index: An Application of the Spectral Second Derivative to Estimate Chlorosis of Leaves in Stressed Vegetation," *International Journal of Remote Sensing*, 20(18):3663–3675.

Adler-Golden, S. M., Matthew, M. W., Anderson, G. P., Felde, G. W., and J. A. Gardner, 2002, "An Algorithm for Deshadowing Spectral Imagery," *Proceedings, Annual JPL AVIRIS Conference*, Pasadena: NASA JPL, 8 p.

Agrawal, G., and J. Sarup, 2011, "Comparison of QUAC and FLAASH Atmospheric Correction Modules on EO-1 Hyperion Data of Sanchi," *International Journal of Advanced Engineering Sciences and Technologies*, 4(1):178–186.

Alonzo, M., Bookhagen, B., and D. A. Roberts, 2014, "Urban Tree Species Mapping Using Hyperspectral and Lidar Data Fusion," *Remote Sensing of Environment*, 148:70–83.

ASTER Spectral Library, 2014, contains the *Johns Hopkins University Spectral Library, Jet Propulsion Laboratory Spectral Library,* and the *U.S. Geological Survey Spectral Library*, Pasadena: NASA JPL, http://speclib.jpl.nasa.gov/.

Baldridge, A. M., Hook, S. J., Grove, C.I. and G. Rivera, 2009, "The ASTER Spectral Library Version 2.0," *Remote Sensing of Environment*, 113:711–715.

Baranoski, G. V., and J. G. Rokne, 2005, "A Practical Approach for Estimating the Red Edge Position of Plant Leaf Reflectance," *International Journal of Remote Sensing*, 26 (3):503–521.

Behmann, J., Steinrucken, J., and L. Plumer, 2014, "Detection of Early Plant Stress Responses in Hyperspectral Images," *ISPRS Journal of Photogrammetry & Remote Sensing*, 93:98–111.

Belluco, E., Camuffo, M., Ferrari, S., Modense, L., Silvestri, S., Marani, A., and M. Marani, 2006, "Mapping Saltmarsh Vegetation by Multispectral Hyperspectral Remote Sensing," *Remote Sensing of Environment*, 105:54–67.

Ben-Dor, E., Patkin, K., Banin, A., and A. Karnieli, 2002, "Mapping of Several Soil Properties using DAIS-7915 Hyperspectral Scanner Data—A Case Study over Clayey Soils in Israel," *International Journal of Remote Sensing*, 23(60):1043–1062.

Bernstein, L. S., Alder-Golden, S. M., Sundberg, R. L., and A. J. Ratkowski, 2008, "In-scene-based Atmospheric Correction of Uncalibrated VISible-SWIR (VIS-SWIR) Hyper- and Multispectral Imagery," *Proceedings, Europe Security and Defense, Remote Sensing*, Volume 7107, 8 p.

Blonski, S., Gasser, G., Russell, J., Ryan, R., Terrie, G., and V. Zanoni, 2002, "Synthesis of Multispectral Bands from Hyperspectral Data: Validation Based on Images Acquired by AVIRIS, Hyperion, ALI, and ETM^+," *Proceedings, Annual AVIRIS Conference*, Pasadena: NASA JPL, 9 p.

Boardman, J. W., 1993, "Automating Spectral Unmixing of AVIRIS Data Using Convex Geometry Concepts," *Summaries of the 4th Annual JPL Airborne Geoscience Workshop*, Pasadena: NASA JPL, Publication 93-26, 1:11–14.

Boardman, J. W., 1997, "Mineralogic and Geochemical Mapping at Virginia City, Nevada, Using 1995 AVIRIS Data," *Proceedings of the 12th Thematic Conference on Geological Remote Sensing*, Ann Arbor: ERIM, 21–28.

Boardman, J. W., 1998, "Post-ATREM Polishing of AVIRIS Apparent Reflectance Data Using EFFORT: A Lesson in

Accuracy Versus Precision," *Summaries of the 7th JPL Airborne Earth Science Workshop*, Pasadena: NASA JPL, 1:53.

Boardman, J. W., and F. A. Kruse, 1994, "Automated Spectral Analysis: A Geological Example Using AVIRIS Data, North Grapevine Mountains, Nevada," *Proceedings, 10th Thematic Conference on Geologic Remote Sensing,* Ann Arbor: ERIM, Vol I: 407–418.

Brantley, S. T., Zinnert, J. C., and D. R. Young, 2011, "Application of Hyperspectral Vegetation Indices to Detect Variations in High Leaf Area Index Temperate Shrub Thicket Canopies," *Remote Sensing of Environment*, 115:514–523.

Brown, L., Chen, J. M., Leblanc, S. G., and J. Cihlar, 2000, "A Shortwave Infrared Modification to the Simple Ratio for LAI Retrieval in Boreal Forests: An Image and Model Analysis," *Remote Sensing of Environment*, 71:16–25.

Camps-Valls, G., and L. Bruzzone, 2005, "Kernel-based Methods for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1351–1362.

Chang, C., 2013, *Hyperspectal Data Processing: Algorithm Design and Analysis*, New York: John Wiley, 1164 p.

Chen, C. M., Hepner, G. F., and R. R. Forster, 2003, "Fusion of Hyperspectral and Radar Data Using the IHS Transformation to Enhance Urban Surface Features," *ISPRS Journal of Photogrammetry & Remote Sensing*, 58:19–30.

Chen, J. M., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S. G., White, H. P., Hall, R. J., Peddle, D. R., King, D. J., Trofymow, J. A., Swift, E., Van der Sanden, J., and P. K. Pellikka, 2002, "Derivation and Validation of Canada-wide Coarse-resolution Leaf Area Index Maps Using High-resolution Satellite Imagery and Ground Measurements," *Remote Sensing of Environment*, 80:165–184.

Chiu, L. S., Zhao, X., Resimini, R. G., Sun, D., Stefanidis, A., Qu, J., and R. Yang, 2011, "Chapter 2: Earth Observations," in Yang, C., Wong, D., Miao, Q., and R. Yang (Eds.), *Advanced Geoinformation Science*, Boca Raton: CRC Press, 17–78.

Clark, M. L., Roberts, D. A. Ewel, J. J., and D. B. Clark, 2011, "Estimation of Tropical Rain Forest Aboveground Biomass with Small-footprint Lidar and Hyperspectral Sensors," *Remote Sensing of Environment*, 115:2931–2942.

Clevers, J. G., 1994, "Imaging Spectrometry in Agriculture: Plant Vitality and Yield Indicators," in Hill, J. and J. Megier (Eds.), *Imaging Spectrometry: A Tool for Environmental Observations*, Dordrecht: Kluwer Academic, 193–219.

Coleman, M.D., et al., 2004, "Production of Short-rotation Woody Crops Grown with a Range of Nutrient and Water Availability: Establishment Report and First-year Responses," *Technical Report SRS-72*, Forest Service, USDA.

Dalponte, M., Bruzzone, L., Vescovo, L., and D. Gianelle, 2009, "The Role of Spectral Resolution and Classifier Complexity in the Analysis of Hyperspectral Images of Forest Areas," *Remote Sensing of Environment*, 113:2345–2355.

Dash, J., Jeganathan, C., and P. M. Atkinson, 2010, "The Use of MERIS Terrestrial Chlorophyll Index to Study Spatio-Temporal Variation in Vegetation Phenology over India," *Remote Sensing of Environment*, 114:1388–1402.

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., Brown de Colstoun, E., and J. E. McMurtrey III, 2000, "Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance," *Remote Sensing of Environment*, 74:229–239.

Dawson, T. P., and P. J. Curran, 1998, "A New Technique for Interpolating the Reflectance Red Edge Position," *International Journal of Remote Sensing*, 19(11):2133–2139.

Demetriades-Shah, T. H., Steven, M. D., and J. A. Clark, 1990, "High Resolution Derivative Spectra in Remote Sensing," *Remote Sensing of Environment*, 33:55–64.

DLR ROSIS, 2014, *ROSIS Hyperspectral System*, Germany: DLR, http://messtec.dlr.de/en/technology/dlr-remote-sensing-technology-institute/hyperspectral-systems-airborne-rosis-hyspex/.

Eismann, M., 2012, *Hyperspectral Remote Sensing,* Monograph Vol. PM210, New York: SPIE Press, 748 p.

Elvidge, C. D., and Z. Chen, 1995, "Comparison of Broad-band and Narrow-band Red and Near-infrared Vegetation Indices," *Remote Sensing of Environment*, 54(1):38–48.

ERDAS ATCOR, 2014, *ATCOR for ERDAS Imaging - Atmospheric Correction for Professionals*, http://www.geosystems.de/atcor/index.html.

ESA MERIS, 2014, *MERIS*, European Space Agency, https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/envisat/instruments/meris.

Exelis ENVI, 2014, *ENVI*, Boulder, Exelis, Inc., http://www.exelisvis.com/ProductsServices/ENVIProducts.aspx.

Exelis FLAASH, 2014, *ENVI Classic Tutorial: Atmospherically Correcting Hyperspectral Data Using FLAASH*, Boulder: Exelis, Inc., 8 p., http://www.exelisvis.com/portals/0/pdfs/envi/FLAASH_Hyperspectral.pdf.

Exelis QUAC, 2014, *QUick Atmospheric Correction (QUAC)*, Boulder: Exelis, Inc., http://www.exelisvis.com/docs/QUAC.html and http://www.exelisvis.com/docs/BackgroundQUAC.html.

Filippi, A. M., and J. R. Jensen, 2007, "Effect of Continuum Removal on Hyperspectral Coastal Vegetation Classification using Fuzzy Learning Vector Quantizer," *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1857–1869.

Franke, J., Roberts, D. A., Halligan, K., and G. Menz, 2009, "Hierarchical Multiple Endmember Spectral Mixture Analysis (MESMA) of Hyperspectral Imagery for Urban Environments," *Remote Sensing of Environment*, 113:1712–1723.

GAC, 2013, *Airborne Imaging Spectroscopy at DLR*, German Aerospace Center, http://www.dlr.de/dlr/en/desktop default.aspx/tabid-10002/.

Gamon, J., Penuelas, J., and C. Field, 1992, "A Narrow-waveband Spectral Index that Tracks Diurnal Changes in photosynthetic Efficiency," *Remote Sensing of Environment*, 41, 35–44.

Gao, B. C., 1996, "NDWI: A Normalized Difference Water Index for Remote Sensing of Liquid Water from Space," *Remote Sensing of Environment,* 58:257–266.

Gao, B. C., Heidebrecht, K. B., and A. F. H. Goetz, 1993, Derivation of Scaled Surface Reflectances from AVIRIS Data," *Remote Sensing of Environment*, 44:165–178.

Gao, B. C., Heidebrecht, K. B., and A. F. H. Goetz, 1999, *ATmosphere REMoval Program (ATREM) User's Guide*, Boulder: Center for Study of the Earth from Space, 31 p.

Gao, B. C., Montes, M. J., Davis, C. O., and A. F. H. Goetz, 2009, "Atmospheric Correction Algorithms for Hyperspectral Remote Sensing of Land and Ocean," *Remote Sensing of Environment*, 113:S17–S24.

Gitelson, A. A., Viæa, A., Ciganda, C., Rundquist, D. C., and T. J. Arkebauer, 2005, "Remote Estimation of Canopy Chlorophyll Content in Crops," *Geophysical Research Letters*, 32, L08403.

Goel, P. K., Prasher, S. O., Patel, R. M., Landry, J. A., Bonnell, R. B., and A. A. Viau, 2003, "Classification of Hyperspectral Data by Decision Trees and Artificial Neural Networks to Identify Weed Stress and Nitrogen Status of Corn," *Computers and Electronics in Agriculture*, 39:67–93.

Goetz, A. F., Vane, G., Solomon, J. E., and B. N. Rock, 1985, "Imaging Spectrometry for Earth Remote Sensing," *Science*, 228: 1147–1153.

Gray, J., and C. Song, 2012, "Mapping Leaf Area Index using Spatial, Spectral, and Temporal Information from Multiple Sensors," *Remote Sensing of Environment*, 119:173–183.

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., and L. Dextraze, 2002, "Integrated Narrow-band Vegetation Indices for Prediction of Crop Chlorophyll Content for Application to Precision Agriculture," *Remote Sensing of Environment*, 81:416–426.

Hatala, J. A., Crabtree, R. L., Halligan, K. Q., and P. R. Moorcroft, 2010, "Landscape-scale Patterns of Forest Pest and Pathogen Damage in the Greater Yellowstone Ecosystem," *Remote Sensing of Environment*, 114:375–384.

Hildebrand, F. B., 1956, *Introduction to Numerical Analysis*, New York: McGraw-Hill, 511 p.

Huang, C., and J. A. Townshend, 2003, "Stepwise Regression Tree for Nonlinear Approximation: Applications to Estimating Subpixel Land Cover," *International Journal of Remote Sensing*, 24:75–90.

Huete, A. R., 1988, "A Soil-adjusted Vegetation Index (SAVI)," *Remote Sensing of Environment*, 25:295–309.

HyVista HyMap, 2014, *Hymap*, Australia: HyVista, Inc., http://www.hyvista.com/?page_id=440.

Im, J., and J. R. Jensen, 2008, "Hyperspectral Remote Sensing of Vegetation," *Geography Compass*, 2:1943–1961.

Im, J., Jensen, J. R., Coleman, M., and E. Nelson, 2009, "Hyperspectral Remote Sensing Analysis of Short Rotation Woody Crops Grown with Controlled Nutrient and Irrigation Treatments," *Geocarto International*, 24(4):293–312.

Im, J., Jensen, J. R., Jensen, R. R., Gladden, J., Waugh, J., and M. Serrato, 2012, "Vegetation Cover Analysis of Hazardous Waste Sites in Utah and Arizona using Hyperspectral Remote Sensing," *Remote Sensing*, 4:327–353.

ImSpec ACORN, 2014, *ACORN 6.1x,* Boulder: ImSpec, LLC, http://www.imspec.com/.

ITRES CASI-1500, 2014, *CASI-1500*, Canada: ITRES Research Ltd., www.itres.com.

ITRES MASI-600, 2014, *MASI-600*, Canada: ITRES Research Ltd., www.itres.com.

ITRES SASI-600, 2014, *SASI-600*, Canada: ITRES Research Ltd., www.itres.com.

ITRES TASI-600, 2014, *TASI-600*, Canada: ITRES Research Ltd., www.itres.com.

Jay, S., and M. Guillaume, 2014, "A Novel Maximum Likelihood Based Method for Mapping Depth and Water Quality from Hyperspectral Remote Sensing Data," *Remote Sensing of Environment*, 147:121–132.

Jensen, J. R., 2007, *Remote Sensing of the Environment: An Earth Resource Perspective*, Boston: Pearson, 592 p.

Jensen, J. R., and R. R. Jensen, 2013, *Introductory Geographic Information Systems*, Boston: Pearson, Inc., 400 p.

Jensen, J. R., Im, J., Hardin, P., and R. R. Jensen, 2009, "Chapter 19: Image Classification," in *The Sage Handbook of Remote Sensing*, Warner, T. A., Nellis, M. D. and G. M. Foody, (Eds.), LA: Sage Publications, 269–296.

Jensen, J. R., and B. Hadley, 2003, *Remote Sensing Analysis of Closure Caps at Savannah River Site and Other Department of Energy Facilities*, Westinghouse Savannah River Company, Aiken. S. C., 80 p.

Jensen, R. R., Hardin, A. J., Hardin, P. J., and J. R. Jensen, 2011, "A New Method to Correct Push-broom Hyperspectral Data using Linear Features and Ground Control Points," *GIScience & Remote Sensing*, 48(4):416–431.

Jia, X., and J. A. Richards, 1993, "Binary Coding of Imaging Spectrometry Data for Fast Spectral Matching and Classification," *Remote Sensing of Environment*, 43:47–53.

Kruse, F. A., 1994, "Imaging Spectrometer Data Analysis — A Tutorial," *Proceedings of the International Symposium on Spectral Sensing Research*, June 10–15, San Diego, CA, Volume I, 44–54.

Li, Y., Demetriades-Shah, T. H., Kanemasu, E. T., Shultis, J. K., and K. B. Kirkham, 1993, "Use of Second Derivatives of Canopy Reflectance for Monitoring Prairie Vegetation

over Different Soil Backgrounds," *Remote Sensing of Environment*, 44:81–87.

Malthus, T. J., and A. C. Madeira, 1993, "High Resolution Spectroradiometry: Spectral Reflectance of Field Bean Leaves Infected by *Botrytis fabae*," *Remote Sensing of Environment*, 45:107–116.

Matthews, M. W., 2011, "A Current Review of Empirical Procedures of Remote Sensing in Inland and Near-Coastal Transitional Waters," *International Journal of Remote Sensing*, 32(21):6855–6899.

McCoy, R., 2005, *Field Methods in Remote Sensing*, NY: Guilford, 159 p.

McGwire, K., Minor, T., and L. Fenstermaker, 2000, "Hyperspectral Mixture Modeling for Quantifying Sparse Vegetation Cover in Arid Environments," *Remote Sensing of Environment*, 72:360–374.

Meigs, A. D., Otten, J., and T. Y. Cherezova, 2008, "Ultraspectral Imaging: A New Contribution to Global Virtual Presence," *IEEE A&E Systems Magazine,* (Oct.):11–17.

Mitchell, J. J., Glenn, N. F., Sankey, T. T., Derryberry, D. R., and M. J. Germino, 2012, "Remote Sensing of Sagebrush Canopy Nitrogen," *Remote Sensing of Environment*, 124:217–223.

Moisen, G., Freeman, E., Blackard, J., Frescino, T., Zimmermann, N., and T. Edwards, 2006, "Predicting Tree Species Presence and Basal Area in Utah: Comparison of Stochastic Gradient Boosting, Generalized Additive Models, and Tree-based Methods," *Ecological Modeling*, 199:176–187.

Mustard, J. F., Staid, M. I., and W. J. Fripp, 2001, "A Semianalytical Approach to the Calibration of AVIRIS Data to Reflectance over Water Application in a Temperate Estuary," *Remote Sensing of Environment*, 75:335–349.

NASA HyspIRI, 2014, *HyspIRI Mission Study*, Pasadena: NASA JPL, http://hyspiri.jpl.nasa.gov/.

Nolin, A. W., and J. Dozier, 2000, "A Hyperspectral Method for Remotely Sensing the Grain Size of Snow," *Remote Sensing of Environment*, 74:207–216.

Okin, G. S., Roberts, D. A., Muray, B., and W. J. Okin, 2001, "Practical Limits on Hyperspectral Vegetation Discrimination in Arid and Semiarid Environments," *Remote Sensing of Environment*, 77:212–225.

Penuelas, J., Gamon, J. A., Griffin, K. L., and C. B. Field, 1993, "Assessing Community Type, Plant Biomass, Pigment Composition, and Photosynthetic Efficiency of Aquatic Vegetation from Spectral Reflectance," *Remote Sensing of Environment*, 46:110–118.

Philpot, W. D., 1991, "The Derivative Ratio Algorithm: Avoiding Atmospheric Effects in Remote Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, 29(3):350–357.

Pignatti, S., Cavalli, R. M., Cuomo, V., Fusillli, L., Pascucci, S., Poscolieri, M., and F. Santini, 2009, "Evaluating Hyperion Capability for Land Cover Mapping in a Fragmented Ecosystem: Pollino National Park, Italy," *Remote Sensing of Environment*, 113:622–634.

Plaza A., and 12 co-authors, 2009, "Recent Advances in Techniques for Hyperspectral Imaging Processing," *Remote Sensing of Environment*, 113:S110–S122.

Prost, G. L., 2013, *Remote Sensing for Geoscientists: Image Analysis and Integration*, 3rd Ed., Boca Raton: CRC Press, 702 p.

Ramsey, M. S., Realmuto, V. J., Hulley, G. C., and S. J. Hook, 2012, *HyspIRI Thermal Infrared (TIR) Band Study Report*, JPL Publication 12–16, Pasadena: NASA JPL, 49 p.

Research Systems, Inc., 2004, *ENVI User's Guide*, Version 4.1, Boulder: Research Systems, Inc., 1150 p.

Richards, J. A. 2013, *Remote Sensing Digital Image Analysis*, 5th Ed., NY: Springer-Verlag, 494 p.

Richter, R., and D. Schlaper, 2012, *Atmospheric / Topographic Correction for Satellite Imagery: ATCOR-2/3 user Guide, Ver. 8.2*, Switzerland: ReSe Applications Schlaper, 216 p.

Roberts, D. A., Smith, M. O., and J. B. Adams, 1993, "Green Vegetation, Nonphotosynthetic Vegetation and Soils in AVIRIS Data," *Remote Sensing of Environment*, 44:255–269.

Rodger, A., Laukamp, C., Haest, M., and T. Cudahy, 2012, "A Simple Quadratic Method of Absorption Feature Wavelength Estimation in Continuum Removed Spectra," *Remote Sensing of Environment*, 118:273–283.

Rondeaux, G., Steven, M., and F. Baret, 1996, "Optimization of Soil-adjusted Vegetation Indices," *Remote Sensing of Environment*, 55:95–107.

Roth, K.L., Dennison, P. E., and D. A. Roberts, 2012, "Comparing Endmember Selection Techniques for Accurate Mapping of Plant Species and Land Cover using Imaging Spectrometer Data," *Remote Sensing of Environment*, 127:139–152.

San, B. T., and M. L. Suzen, 2010, "Evaluation of Different Atmospheric Correction Algorithms for EO-1 Hyperion Imagery," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, 38(8):392–397.

Schaepman, M. E., Ustin, S. L., Plaza, A. J., Painter, T. H., Verrelst, J., and S. Liang, 2009, "Earth System Science Related Imaging Spectroscopy – An Assessment," *Remote Sensing of Environment*, 113:123–137.

Schowengerdt, R. A., 2007, *Remote Sensing: Models and Methods for Image Processing*, NY: Academic Press, 522 p.

Schuerger, A. C., 2003, unpublished, "Use of Laser-induced Fluorescence Spectroscopy and Hyperspectral Imaging to Detect Plant Stress in Bahiagrass Grown under Different Concentrations of Zinc and Copper."

Segl, K., Roessner, S., Heiden, U., and H. Kaufmann, 2003, "Fusion of Spectral and Shape Features for Identification of Urban Surface Cover Types Using Reflective and Ther-

mal Hyperspectral Data," *ISPRS Journal of Photogrammetry & Remote Sensing*, 58:99–112.

Small, C., 2012, "Spatiotemporal Dimensionality and Time-Space Characterization of Multitemporal Imagery," *Remote Sensing of Environment*, 14:793–809.

Smith, K. L., Steven, M. D., and J. J. Colls, 2004, "Use of Hyperspectral Derivative Ratios in the Red-edge Region to Identify Plant Stress Responses to Gas Leaks," *Remote Sensing of Environment*, 92:207–217.

Somers, B., Asner, G. P., Tits, L., and P. Coppin, 2011, "Endmember Variability in Spectral Mixture Analysis: A REview," *Remote Sensing of Environment*, 115:1603–1616.

Specim AISA, 2014, *Aisa Airborne Hyperspectral Sensors*, Oulu, Finland: Spectral Imaging Ltd., http://www.specim.fi/index.php/products/airborne.

Stagakis, S., Markos, N., Sykioti, O., and A. Kyparissis, 2010, "Monitoring Canopy Biophysical and Biochemical Parameters in Ecosystem Scale using Satellite Hyperspectral Imagery: Application on a Phlomis fruticosa Mediterranean Ecosystem using Multiangular CHRIS/PROBA Observations," *Remote Sensing of Environment*, 114:977–994.

Talsky, G., 1994, *Derivative Spectrophotometry: Low and Higher Order*, New York: VCH Publishers, 228 p.

Thenkabail, P. S., Enclona, E. A., Ashston, M. S., and B. Van der Meer, 2004, "Accuracy Assessments of Hyperspectral Waveband Performance for Vegetation Analysis Applications," *Remote Sensing of Environment,* 91:354–376.

Thenkabail, P. S., Lyon, J. G., and A. Huete, 2011, *Hyperspectral Remote Sensing of Vegetation*, Boca Raton: CRC Press, 781 p.

Thenkabail, P. S., Smith, R. B., and E. De Paww, 2000, "Hyperspectral Vegetation Indices and Their Relationships with Agricultural Crop Characteristics," *Remote Sensing of Environment,* 71:158–182.

Tsai, F., and W. Philpot, 1998, "Derivative Analysis of Hyperspectral Data," *Remote Sensing of Environment*, 66:41–51.

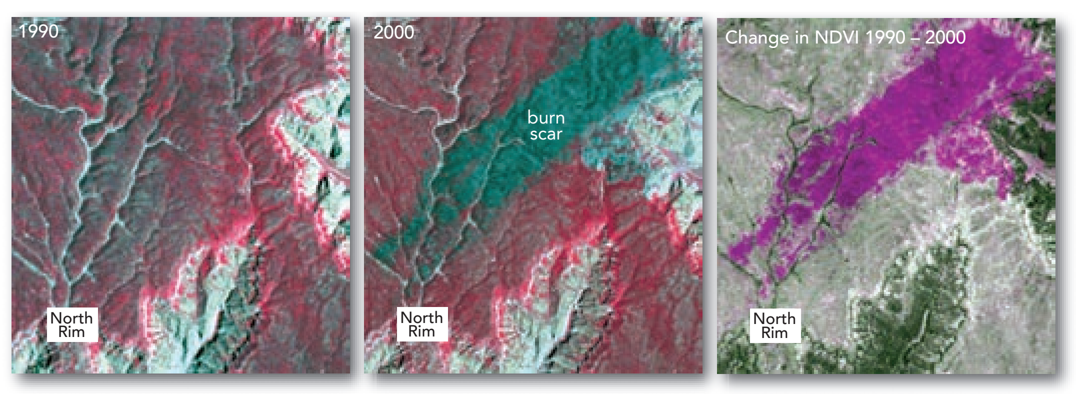
Underwood, E., Ustin, S., and D. DiPietro, 2003, "Mapping Nonnative Plants Using Hyperspectral Imagery," *Remote Sensing of Environment*, 86:150–161.

Ustin, S. L., et al., 2009, "Retrieval of Foilar Information About Plant Pigment Systems from High Resolution Spectroscopy," *Remote Sensing of Environment*, 113:67–77.

Williams, A. P., and E. R. Hunt, 2002, "Estimation of Leafy Spurge Cover from Hyperspectral Imagery Using Mixture Tuned Matched Filtering," *Remote Sensing of Environment*, 82:446–456.

Youngentob, K. N., Roberts, D., Held, A., Dennison, P., Jia, X., and D. Lindenmayer, 2011, "Mapping Two *Eucalyptus* Subgenera using Multiple Endmember Spectral Mixture Analysis and Continuum-removed Imaging Spectrometry Data," *Remote Sensing of Environment*, 115:1115–1128.

12 CHANGE DETECTION



Source: USGS and NASA

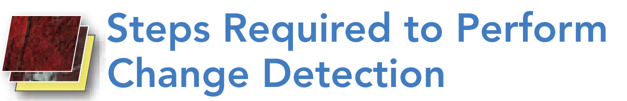
Biophysical materials and human-made features on the surface of Earth are inventoried using remote sensing and *in situ* techniques. Some of the data are fairly static; they change very little over time. Conversely, some biophysical materials and human-made features are dynamic, changing rapidly. It is important that such changes be inventoried accurately so that the physical and human processes at work can be more fully understood (e.g., Miller et al., 2003; Leprince et al., 2007; Jensen et al., 2012; Kit and Ludeke, 2013). It is not surprising, therefore, that significant effort has gone into the development of remote sensing change detection methods (e.g., Patra et al., 2013; Zhu and Woodcock, 2014; Rokni et al., 2015).

Land-use/land-cover change is a major variable impacting global climate change (Foley et al., 2005). This is one reason that the Multi-Resolution Land Characteristics (MRLC) partners who guide the creation of the U.S. National Land Cover Database (NLCD) products have shifted the emphasis from characterizing land cover every ten years to monitoring land cover change every five years (Fry et al., 2011).



Overview

This chapter reviews how change information is extracted from remotely sensed data using digital image processing techniques. It summarizes the remote sensor system and environmental parameters that must be considered when conducting change detection. Several of the most widely used change detection algorithms are introduced and demonstrated.



Steps Required to Perform Change Detection

The general steps required to perform digital change detection using remotely sensed data are summarized in Figure 12-1. It is instructive to review the characteristics of these steps.

Specify the Thematic Attribute(s) or Indicator(s) of Interest

Clear identification of the attributes or indicators of interest is required. Perhaps you are interested in documenting the change in vegetation species distribution through time or monitoring urban expansion into rural areas. Perhaps you want to immediately monitor the location of areas inundated by floodwaters or the location of affected trees as a disease migrates through a forest. Common human-made and natural resource attributes or indicators that are often the focus of land-use planning and natural resource monitoring are summarized in Table 12-1 (Kennedy et al., 2009).

Specify the Change Detection Geographic Region of Interest (ROI)

In conjunction with the specification of the thematic attribute or indicator of interest, the dimensions of the change detection region of interest (ROI) must be carefully identified and held constant throughout a change detection project. The geographic ROI (e.g., a county, state, region or watershed) is especially important in a change detection study because it must be completely covered by *n* dates of imagery. Failure to ensure that each of the multiple-date images covers the geographic

TABLE 12–1 Common human-made and natural resource attributes or indicators that are often the focus of land-use planning and natural resource monitoring programs (expanded from Kennedy et al., 2009).

Man-made and/or Natural Resource Attributes or Indicators	Process of Interest or Threat
Change in the size or shape of areal patches (polygons) of related cover types	Urban expansion (sprawl); vegetation expansion; sea-level rise; consolidation, fragmentation; infilling; encroachment; erosion/dilution
Change in width or character of linear features	Densification of road, utility, or hydrologic network; impact of use of paths or roads; impact of flooding on riparian vegetation; dynamics of terrestrial and submerged near-shore aquatic vegetation
Slow changes in surface cover types or species composition	Succession, competition, eutrophication, consolidation, fragmentation, exotic species invasion
Abrupt changes in surface cover, water, and/or atmospheric condition	Catastrophic event (e.g., hurricane, flood, tornado, volcanic eruption, fire, wind, landslides), disturbance, human activity (e.g., land clearing; urban and/or environmental terrorism), land management practices (e.g., no-till farming; prescribed burning)
Slow changes in condition of a single cover type	Climate-related changes in vegetation species composition and/or productivity; sea surface temperature; slowly-spreading forest mortality caused by insect or diseases; changes in moisture regime
Changes in timing or the extent of diurnal and/or seasonal processes	Coastal zone dynamics, snow cover dynamics, natural vegetation and agriculture phenology

area of interest results in change detection maps with *data voids* that are problematic when computing change statistics.

Specify the Change Detection Time Period

Sometimes change detection studies are overly ambitious in their attempt to monitor changes in the landscape. Sometimes the time period selected over which change is to be monitored is too short or too long to capture the information of interest. Therefore, the analyst must be careful to identify the optimal change detection time period(s). This selection, of course, is dictated by the nature of the problem. Traffic transportation studies might require a change detection period of just a few seconds or minutes. Conversely, images obtained monthly or seasonally might be sufficient to monitor the greening-up of a continent. Careful selection of the change detection time period can ensure that resource analysis funds are not wasted.

It is important to note that just because multiple-date remote sensing imagery is available does not ensure that change detection will be successful. The multiple-date imagery may have been acquired for an entirely different purpose. The most effective change detection projects are those based on timely multiple-date imagery selected specifically for the project that meet the system, environmental, and bureaucratic requirements to be discussed.

Select an Appropriate Land-Use/ Land-Cover Classification System

As discussed in Chapter 9, it is wise to use an established, standardized land-use/land-cover classification system for change detection, such as the following:

- American Planning Association *Land-Based Classification Standard* (LBCS);
- U.S. Geological Survey *Land Use/Land Cover Classification System for Use with Remote Sensor Data*;
- U.S. National Land Cover Dataset (NLCD) Classification scheme;
- NOAA Coastal Change Analysis Program (C-CAP) classification scheme;
- U.S. Department of the Interior Fish & Wildlife Service, *Classification of Wetlands and Deepwater Habitats of the United States;*
- U.S. *National Vegetation & Classification Standard*;
- International Geosphere-Biosphere Program *Land Cover Classification System*.

The use of a standardized classification system allows change information derived from one project to be compared with change derived from another project.

Select Hard (Crisp) and/or Soft (Fuzzy) Change Detection Logic

Most change detection studies have been based on the comparison of multiple-date *hard* (crisp) land-cover classifications of remotely sensed data. The result is the

General Steps Used to Conduct Digital Change Detection Using Remote Sensor Data

State the nature of the change detection problem.

- Specify thematic attribute(s) or indicator(s) of interest.
- Specify change detection geographic region of interest (ROI).
- Specify change detection time period (e.g., daily, seasonal, yearly).
- Define the classes of interest in a classification system.
- Select hard and/or fuzzy change detection logic.
- Select per-pixel or object-based (OBIA) change detection.

Considerations of significance when performing change detection.

- Remote sensing system considerations:
  - Spatial, spectral, temporal, and radiometric resolution
- Environmental considerations:
  - Atmospheric conditions
    - Cloud cover and cloud shadow
    - Relative humidity
  - Soil moisture conditions
  - Phenological cycle characteristics
    - natural (e.g., vegetation, soil, water, snow and ice)
    - man-made phenomena
  - Obscuration considerations
    - natural (e.g., trees, shadows)
    - man-made (e.g., structures, shadows)
  - Tidal stage, etc.

Process remote sensor data to extract change information.

- Acquire appropriate change detection data:
  - *In situ* ground reference information
  - Collateral data (e.g., soil maps, parcels, digital elevation model)
  - Remotely sensed data:
    - Base year (time n)
    - Subsequent year(s) (time $n - 1$ or $n + 1$)
- Preprocess the multiple-date remote sensor data:
  - Geometric correction
  - Radiometric correction (or normalization) if necessary
- Select change detection algorithm.
  - Binary "change versus non-change" algorithm
  - Thematic "from-to" algorithm
- Apply appropriate image classification logic if necessary:
  - Method (supervised, unsupervised, hybrid)
  - Parametric (e.g., maximum-likelihood)
  - Non-parametric (e.g., decision or regression trees, random forests)
- Perform change detection using GIS algorithms if necessary:
  - Highlight selected classes using change detection matrix
  - Generate change-map products
  - Compute change statistics

Perform accuracy assessment.

- Select method:
  - Qualitative confidence building
  - Statistical measurement
- Determine number of samples required by class.
- Select sampling scheme.
- Obtain ground reference test information.
- Create and analyze change detection error matrix:
  - Univariate and multivariate statistical analysis

Accept or reject previously stated hypothesis.
Distribute results if accuracy is acceptable.

FIGURE 12–1 The general steps used to perform digital change detection of remotely sensed data.

creation of a *hard* change detection map consisting of information about the change in discrete categories (e.g., change in forest, agriculture). This is still very important and practical in many instances, but we now recognize that it is sometimes useful to capture both discrete and *soft* (fuzzy) changes in the landscape (refer to Chapter 9 for a discussion about fuzzy land-cover classification).

Land-cover changes may range from no landscape alteration whatsoever, through modifications of variable intensity, to a full transformation or conversion to an entirely new class (e.g., see the Denver, CO, example later in this chapter). Scientists now believe that replacing the Date *n* and Date *n + 1* hard classification maps typically used in a change detection project with fuzzy classification maps can result in more informative and accurate land-cover change information (Foody, 2001; NRC, 2010). However, as discussed in Chapter 13, it is more difficult to determine the accuracy of soft (fuzzy) change detection map products.

Select Per-pixel or Object-Based Change Detection (OBCD)

The majority of digital image change detection has been based on processing Date *n* and Date *n + 1* classification maps pixel by pixel. This is commonly referred to as *per pixel* change detection. Conversely, *object-based change detection* (OBCD) involves the comparison of two or more scenes consisting of many homogenous image objects (i.e., patches) that were identified using the object-based image analysis (OBIA) techniques discussed in Chapter 9 (e.g., Im, 2006; Hofmann et al., 2008; Tsai et al., 2011; Gartner et al., 2014). The homogeneous image objects (i.e., polygons) in the two scenes are then subjected to change detection techniques discussed in this chapter.

Remote Sensing System Change Detection Considerations

Successful remote sensing change detection requires careful attention to:

- remote sensor system considerations, and
- environmental characteristics.

Failure to understand the impact of the various parameters on the change detection process can lead to inaccurate results (Lunetta and Elvidge, 2000; Jensen et al., 2009). Ideally, the remotely sensed data used to perform change detection is acquired by a remote sensor system that holds the following resolutions constant: temporal resolution, look angle, and spatial, spectral, and radiometric resolution. It is instructive to review each of these parameters and identify why they can have a significant impact on the success of a remote sensing change detection project. Kennedy et al. (2009) provide additional guidance on the spatial, spectral, and temporal characteristics of the remote sensor data for a variety of change detection applications.

Temporal Resolution

Two important temporal resolutions should be held constant during change detection using multiple dates of remotely sensed data, if possible. Ideally, the remote sensing data used in the change detection should be obtained from a sensor system that acquires data at approximately the *same time of day*. For example, Landsat Thematic Mapper (TM) data are acquired before 9:45 a.m. for most of the conterminous United States. This eliminates diurnal Sun angle effects that can cause anomalous differences in the reflectance properties of the remotely sensed data.

Whenever possible it is desirable to use remotely sensed data acquired on *anniversary dates*; for example, June 1, 2012, and June 1, 2013. Using anniversary date imagery minimizes the influence of seasonal Sun-angle and plant phenological differences that can negatively impact a change detection project (Dobson et al., 1995; Jensen, 2007; Chen et al., 2012). In addition, all of the shadows found in the anniversary date imagery obtained by the same sensor system at approximately the same time of day have the same shadow orientation and length. This is very important when extracting change information from high spatial resolution imagery because changes in shadow length and orientation can introduce many inaccurate change artifacts into the study (Jensen et al., 2012).

Of course, anniversary date imagery is not practical when monitoring dynamic change associated with forest fires, floods, hurricanes, volcanic eruptions and when monitoring seasonal vegetation (e.g., agriculture) (Chen et al., 2012). It is also important to note that while many change detection projects deal with just two images at a time, there are also studies that utilize dense *time-stacks* of multiple-date imagery in the change detection process (e.g., Lunetta et al., 2006; Schneider, 2012).

It is useful to consider a hypothetical change detection project to demonstrate several variable considerations. For example, suppose it was necessary to map the change in land cover of San Francisco, CA, over a specified time period using high spatial resolution DigitalGlobe QuickBird imagery consisting of both panchromatic and multispectral data collected in 2005, 2007, 2009, and 2011. For bureaucratic reasons it is necessary to only use imagery collected from October 1st through December 15th (bureaucratic constraints are often one of the most difficult criteria to satisfy in a

TABLE 12–2 DigitalGlobe, Inc., QuickBird imagery available for change detection in 2005, 2007, 2009, and 2011 with the following criteria: imagery collected from Oct. 1–Dec. 15; maximum cloud cover (20%); maximum off-Nadir look angle (30°); minimum Sun elevation angle (25°) (information courtesy of DigitalGlobe, Inc.).

Acquisition Date year-mo-day	Maximum Off-Nadir look-angle (°)	Minimum Sun Elevation (°)	Cloud Cover (%)
2005.10.05	24.45	46.58	0
2005.11.05	13.35	35.88	0
2005.11.23	16.72	31.25	0
2007.11.11	26.63	34.28	0
2009.10.16	3.11	41.71	0
2009.11.21	9.48	30.95	0
2011.11.04	24.58	34.03	0
2011.12.03	27.40	27.25	0

change detection project). We would want imagery that has <20% cloud cover, a minimum Sun elevation of 25°, and off-Nadir look angles of <30° (to be discussed). Actual images in the DigitalGlobe, Inc., archive that meet these selected criteria are listed in Table 12-2. Note that because it is late in the year (October–December) in the northern hemisphere that the minimum Sun angles are low as expected but all exceed 25°. Based on these criteria it appears that there are sufficient dates in the archive to conduct the change detection project as all of the images have 0% cloud cover. The analyst would have to select a single image from each date to complete the project after considering look angle.

Look Angle

Some remote sensing systems like GeoEye-1, QuickBird, WorldView-2, and airborne sensors (e.g., Pictometry) collect data at off-Nadir *look angles* as much as ±30°; that is, the sensors obtain data of an area on the ground from an *oblique* vantage point. Imagery used in the change detection process with significantly different look angles can cause problems. For example, consider a maple forest consisting of very large, randomly spaced trees. A QuickBird image acquired at approximately 0° off-Nadir will look directly down on the top of the canopy. Conversely, a QuickBird image acquired at 27° off-Nadir will record reflectance information from the side of the canopy. Differences in reflectance from the two datasets may cause spurious change detection results. Therefore, the data used in a remote sensing digital change detection should be acquired with approximately the same look angle, if possible (Chen et al., 2012). Most change detection algorithms do not work well when the look angle on one date of imagery is dramatically different from the look angle on the other date(s) of imagery. Of course, imagery acquired at nadir is preferable.

A review of the imagery available for the hypothetical San Francisco, CA, change detection study summarized in Table 12-2 reveals that there are significant differences between the maximum off-Nadir look angles. Three of the years (2005, 2007, and 2011) have look angles of >13°. Only the 2009 imagery has look angles <10°. The analyst conducting the change detection study would have to carefully weight the importance of look angle in conjunction with the other variables when selecting the imagery to be analyzed.

Spatial Resolution

Accurate spatial registration of at least two images is essential for digital change detection (Roy, 2000; Kennedy et al., 2009; Klemas, 2011). Ideally, the remotely sensed data are acquired by a sensor system that collects data with the same *instantaneous field-of-view* (IFOV) on each date. For example, Landsat Thematic Mapper data collected at 30 × 30 m spatial resolution on two dates are relatively easy to register to one another. However, it is possible to perform change detection using data collected from two different sensor systems with different IFOVs, for example, Landsat TM data (30 × 30 m) for Date 1 and SPOT HRV XS data (20 × 20 m) for Date 2. In such cases, it is usually necessary to decide on a representative minimum mapping unit (e.g., 20 × 20 m) and then resample both datasets to this uniform pixel size. This does not present a significant problem as long as the image analyst remembers that the information content of the resampled data can never be greater than the IFOV of the original sensor system (i.e., even though the Landsat TM data may be resampled to 20 × 20 m pixels, the information was still acquired at 30 × 30 m resolution and we should not expect to be able to extract additional spatial detail from the TM dataset).

Geometric rectification algorithms (Chapter 6) are used to register the images to a standard map projection or coordinate system (e.g., Universal Transverse Mercator). Rectification should result in the two images having a root mean square error (RMSE) of ≤ 0.5 pixel. Misregistration of the two images may result in the identification of spurious areas of change between the datasets (Klemas, 2011). For example, just one pixel misregistration may cause a stable road on the two dates to show up as a new road in the change image.

Spectral Resolution

A fundamental assumption of digital change detection is that a difference exists in the spectral response of a

pixel on two dates if the biophysical materials within the IFOV have changed between dates. Ideally, the spectral resolution of the remote sensor system is sufficient to record reflected radiant flux in spectral regions that best capture the most descriptive spectral attributes of the object. However, different sensor systems do not record energy in exactly the same regions of the electromagnetic spectrum (i.e., bandwidths). For example, the SPOT 1, 2, and 3 HRV sensors collected data in three relatively broad multispectral bands and one panchromatic band. The Landsat ETM$^+$ collects data in six relatively broad optical bands, one thermal infrared band, and one broad panchromatic band (Chapter 2). Ideally, the same sensor system is used to acquire imagery on multiple dates. When this is not possible, the analyst should *select bands that approximate one another.* For example, SPOT bands 1 (green), 2 (red), and 3 (near-infrared), can be used successfully with Landsat ETM$^+$ bands 2 (green), 3 (red), and 4 (near-infrared). Many of the change detection algorithms do not function well when the spectral characteristics of bands from one sensor system do not match those of another sensor system (Jensen et al., 2012).

Radiometric Resolution

As previously mentioned in Chapter 2, it is possible to digitize analog aerial photography and convert it into digital remote sensor data at a specific spatial resolution (e.g., 1×1 m) and a specific radiometric resolution (e.g., 8-bit values from 0 to 255) and then use the data for change detection purposes. Conversely, most modern digital remote sensing systems routinely collect data at radiometric resolutions ranging from 8-bit (values from 0 to 255) to 11-bit (values from 0 to 2,047). For example, GeoEye-1 and WorldView-2 remote sensing systems collect multispectral data with 11-bit radiometric resolution. Generally, the higher the radiometric resolution, the greater the probability that just noticeable differences in spectral reflectance will be able to be discerned between various objects on the terrain (e.g., between building rooftop materials and surrounding tree and grass vegetation).

Ideally, remote sensor data are collected on multiple dates using the identical remote sensing system, which holds the radiometric resolution variable constant. However, if one of the dates of imagery has a radiometric resolution that is lower than the other (e.g., one consists of 8-bit data and one consists of 10-bit data), then it is general practice to transform the lower resolution data (e.g., 8-bit with values from 0 to 255) to the same radiometric level as the superior radiometric resolution (e.g., 10-bit with values from 0 to 1,023). Of course, the radiometric characteristics of the 8-bit data are not actually improved during this process. But the transformed 8-bit data now have a range of values that is compatible with the 10-bit data.

An even more elegant solution is to convert the 8-bit data to scaled percent reflectance (values from 0 to 100%) and convert the 10-bit data to scaled percent reflectance (values from 0 to 100%). This is common practice when monitoring continuous biophysical variables such as leaf-area-index (LAI) or biomass through time. However, the use of scaled percent reflectance data is not always mandatory when conducting urban image classification and change detection. Similarly, most of the multiple date digital imagery analyzed using soft-copy photogrammetric techniques is not transformed into scaled percent reflectance values prior to processing (Jensen et al., 2012).

Environmental/Developmental Considerations of Importance When Performing Change Detection

Failure to understand the impact of various environmental characteristics on the remote sensing change detection process can lead to inaccurate results. When performing change detection, it is desirable to hold as many environmental variables as constant as possible (Jensen et al., 2009).

Atmospheric Conditions

Ideally, the sky has 0% cloud cover and there is low relative humidity on the dates that imagery is collected for change detection purposes. When atmospheric conditions are not ideal, it may be necessary to a) delete some of the images from further consideration, and/or b) apply various atmospheric corrections to the remote sensor data so that spurious change due to atmospheric conditions is not identified (Jensen et al., 2012).

Clouds: The most accurate image-to-image optical change detection is performed using aerial photography or digital satellite remote sensor data that are acquired with 0% cloud cover (e.g., Figure 12-2a,c of Charleston, SC). Clouds introduce extremely serious problems to change detection studies. Each cloud on a single image obscures the terrain below from investigation (Figure 12-2b,d). It causes serious problems during the change detection because there is no land cover or structural information available where the cloud was present, resulting in ambiguous change detection in this area. All the user can do is create a mask and identify certain areas as incomplete because of "cloud cover problems" (Jensen et al., 2012).

Cloud shadow: In addition, almost every cloud casts a shadow that is visible in the imagery except when the cloud is at Nadir relative to the sensor system. Usually the cloud shadow is some angle away from the actual cloud in the image due to a) sensor viewing geometry conditions (e.g., the sensor is pointed 10° off-Nadir),

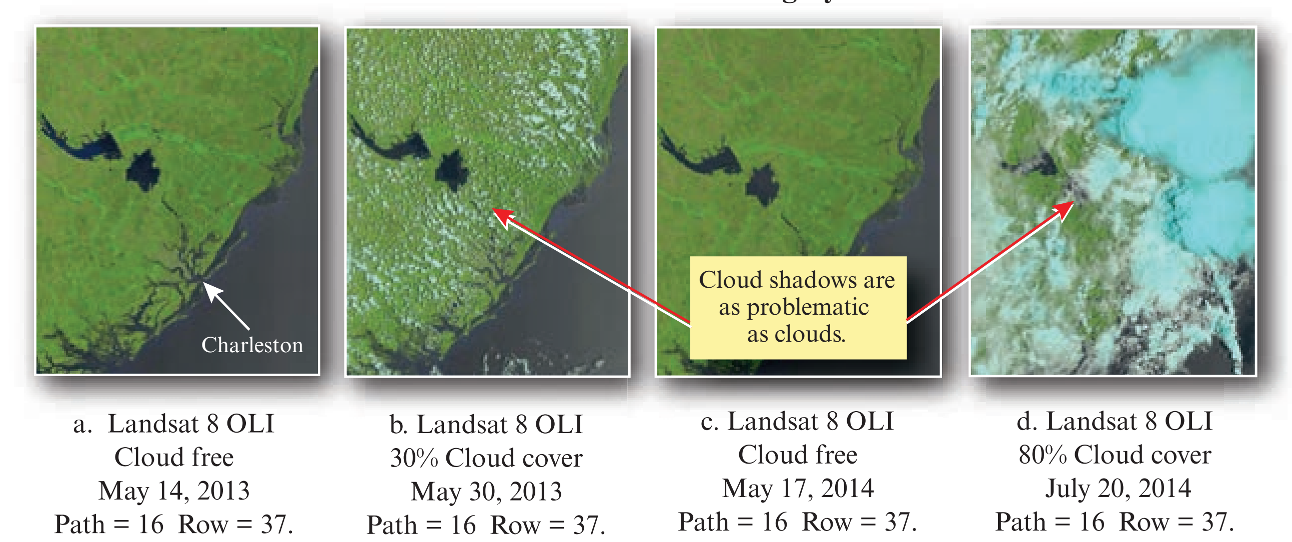


FIGURE 12–2 a,c) Zero percent cloud cover on Landsat 8 scene Path 16/Row 37 on May 14, 2013 and May 17, 2014, respectively. These data are excellent for change detection purposes. b,d) 30% cloud cover on May 30, 2013 and massive 80% cloud cover on July 20, 2014. The use of these images for change detection would yield poor results (imagery courtesy of NASA).

or b) the angle of the Sun which is a function of the time of year (e.g., June 21, the Summer Solstice) and the latitude of the study area (e.g., 30° N. Latitude). The terrain in the shadow area may be visible in the imagery, but it is usually much darker in tone due to the reduced amount of solar illumination reaching the terrain (Figure 12-2b,d). This causes the same feature (e.g., building, road, stand of trees) to have substantially different spectral reflectance characteristics than the same feature observed on a different date (e.g., the next day) using the identical remote sensing data collection parameters with 0% cloud cover. When it is absolutely necessary to extract information from a geographic area that lies in cloud shadow, it is best to stratify this area (i.e., mask it out) and classify it separately using its unique spectral properties. Once classified, the masked out area can be digitally stitched back into the final date dataset and then used in the change detection procedure (Jensen et al., 2012).

Relative Humidity and Other Atmospheric Constituents: Often the imagery used for change detection is free of clouds but there is substantial water vapor in the atmosphere resulting in high relative humidity or in certain areas there may be elevated amounts of haze or smog in the atmosphere. When this occurs, it may be necessary to apply atmospheric correction to the individual dates of remote sensor data prior to change detection. This is done using either a) relative atmospheric radiometric normalization techniques (Jensen et al., 1995; Coulter et al., 2011), or b) absolute radiometric correction discussed in Chapter 6.

Soil Moisture Conditions

Ideally, the soil moisture in the study area is the same (held constant) on both of the images used to detect change. Unfortunately, this is often not the case. Frontal systems often move into an area and drop a relatively consistent amount of precipitation on the landscape. When this occurs, the entire area under investigation might have similar soil moisture conditions. Conversely, sometimes thunderstorms drop precipitation over randomly located areas in a lobe-shaped pattern. This results in randomly located areas that have higher soil moisture than surrounding or adjacent areas. Of course, the absence of soil moisture during drought conditions can also create a landscape that is extremely dry (Jensen et al., 2012).

Geographic areas with extremely low soil moisture typically will exhibit substantially higher spectral reflectance throughout the visible and near-infrared portions of the spectrum than the same terrain when it is saturated and has high soil moisture content. Therefore, multiple-date imagery of terrain that is either very moist or very dry can be problematic when subjected to change detection algorithms, resulting in spurious regions of change being identified due to differences in soil moisture rather than actual changes in features. If the two images under examination exhibit substantial differences in soil moisture, it may be possible to apply radiometric normalization to one of the images to make it have approximately the same spectral reflectance characteristics as the other image (e.g., Jensen et al., 1995; 2009). If only a portion of an image exhibits higher or lower than normal soil moisture conditions

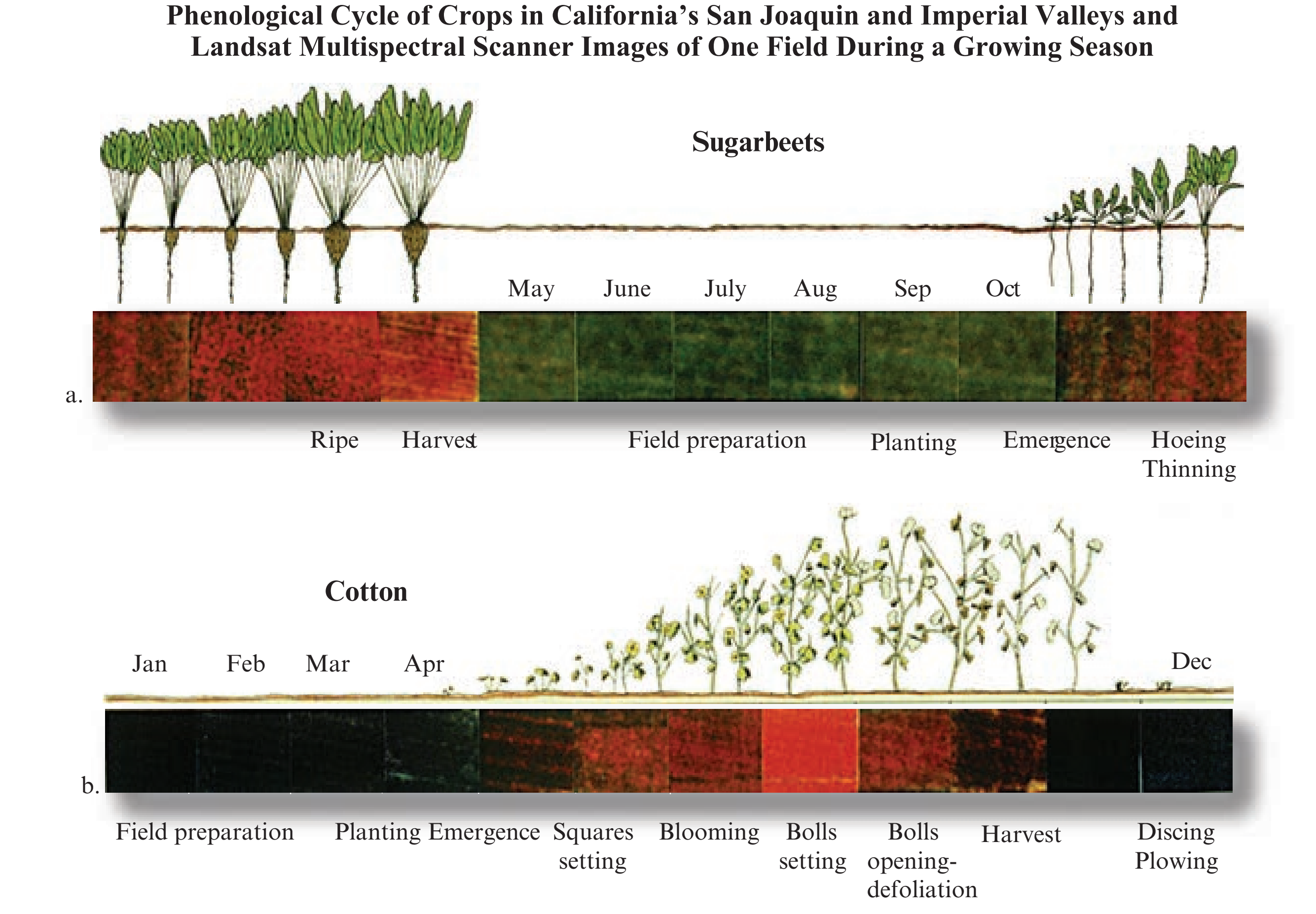


FIGURE 12–3 Phenological cycles of a) sugarbeets, and b) cotton grown in the California San Joaquin Valley. Landsat MSS images were obtained over a 12-month period. The color composite images (MSS RGB = bands 4, 2, 1) of two fields were extracted and placed below the crop calendar information.

due to thunderstorm precipitation or drought conditions, then it is best to stratify (mask-out) this area out and analyze it separately to extract land-cover information. This stratified area is then reinserted into the land-cover classification database prior to performing the change detection.

Phenological Cycle Characteristics

Natural ecosystems go through predictable cycles of development. Human beings often modify the landscape in repeatable, predictable stages. These cycles of predictable development are often referred to as phenomenological or **phenological cycles**. Image analysts use these cycles to identify when remotely sensed data should be collected to obtain the maximum amount of usable change information. Therefore, analysts must be intimately familiar with the *biophysical* characteristics of the vegetation, soils, and water constituents of ecosystems and their phenological cycles. Likewise, it is imperative that they understand the phenological cycles associated with human-made development, such as residential expansion at the urban/rural fringe.

Vegetation Phenology: Healthy vegetation grows according to relatively predictable diurnal, seasonal, and annual phenological cycles. Using approximately near-anniversary images greatly minimizes the effects of seasonal phenological differences that may cause spurious change to be detected in the imagery. For example, when attempting to identify change in agricultural crops using remote sensor data, the analyst must be aware of when the ground is prepared, when the crops are planted, when they reach maturity, and when they are typically harvested. Ideally, monoculture crops (e.g., corn, wheat, sugarbeets, cotton) *are planted at approximately the same time of year*. A month lag in planting date between fields of the same crop can cause serious change detection error. Second, the monoculture crops should be *the same species.* Different species of a crop can cause the crop to reflect energy differently on the multiple dates of anniversary imagery. In addition, changes in row spacing and direction can have an impact. These observations suggest that the analyst must know the crop's *biophysical* characteristics as well as the *cultural* land-tenure practices in the study area

Phenological Cycle of Cattails and Waterlilies in Par Pond in South Carolina

Relative Greenness
Cattails are leafed-out and green
Cattails are senesced but persist from year to year
Maximum cloud cover and humidity
Waterlilies are leafed-out and green
Best temporal window for discriminating between cattails and waterlilies
It is difﬁcult to distinguish between cattails and waterlilies from June through August because they have similar spectral characteristics
Cattails senesce
Minimum cloud cover and lower humidity
Waterlilies die
Relative Growth
J F M A M J J A S O N D
Month

FIGURE 12–4 Annual phenological cycle of cattails and waterlilies in Par Pond, SC. Cattails are present throughout the year while waterlilies emerge in April and disappear in November. The best time to acquire imagery to discriminate between cattails and waterlilies is from mid-April to mid-May.

so that the most appropriate remotely sensed data can be selected for change detection.

For example, consider the phenological characteristics of two of the most important crops grown in the San Joaquin Valley in California shown in Figure 12-3. If we wanted to monitor the change in sugarbeet production during a growing season then we would have to acquire imagery from September through April. If we wanted to detect the change in sugarbeet production in multiple years, it would be wise to evaluate imagery collected from September through April. Conversely, if we wanted to monitor the change in cotton production through the season or in multiple years, we would analyze imagery collected from May through October.

Natural vegetation ecosystems such as wetland, forests, and rangeland have unique phenological cycles. For example, consider the phenological cycle of cattails and waterlilies found in lakes in the southeastern United States (Figure 12-4). Cattails persist year round in lakes and are generally found in shallow water adjacent to the shore (Jensen et al., 1993b). They begin greening up in early April and often have a full, green canopy by late May. Cattails senesce in late September to early October, yet they are physically present and appear brown through the winter months. Conversely, waterlilies and other nonpersistent species do not live through the winter. Waterlilies appear at the outermost edge of the cattails in mid-April and reach full emergence 6 to 8 weeks later. The waterlily beds usually persist above water until early November, at which time they die.

The phenological cycles of cattails and waterlilies dictate the most appropriate times for remote sensing data acquisition. It is generally difficult to discriminate between cattails and waterlilies from June through August because their spectral characteristics are very similar. This is also the time of maximum cloud cover and high humidity which reduces the availability of useful imagery. Conversely, from mid-April through mid-May, the cattails are well-developed while the waterlilies are just beginning to become established. This results in spectral reflectance differences between cattails and waterlilies as suggested by the red arrow in Figure 12-4.

Many remote sensing change detection studies have failed because the analyst did not use accurate vegetation phenological information when selecting the remote sensing imagery to be analyzed.

Urban–Suburban Phenological Cycles: Urban and suburban landscapes also tend to have predictable phenological cycles. For example, consider the residential

Residential Development near Denver, CO, from October 8, 1976, to October 15, 1978



a. Panchromatic photograph obtained on October 8, 1976.



b. Panchromatic photograph obtained on October 15, 1978.

FIGURE 12–5 a) Panchromatic aerial photograph of a portion of the Fitzsimmons 7.5-minute quadrangle near Denver, CO, on October 8, 1976. The original scale of the photography was 1:52,800. The land cover was visually photo-interpreted and classified into 10 classes of residential development using the logic shown in Figure 12-6. b) Panchromatic aerial photograph of a portion of the Fitzsimmons 7.5-minute quadrangle on October 15, 1978. The original scale of the photography was 1:57,600. Comparison with the 1976 aerial photography reveals substantial residential land development.

development from 1976 to 1978 in the 6-mi^2 portion of the Fitzsimmons 7.5-minute quadrangle near Denver, CO. Aerial photographs obtained on October 8, 1976, and October 15, 1978, reveal dramatic changes in the landscape (Figure 12-5). Most novice image analysts assume that change detection in the urban–rural fringe will capture the residential development in the two most important stages: rural undeveloped land and completely developed residential. Jensen (1981) identified 10 stages of residential development taking place in this region based on evidence of clearing, subdivision, transportation, buildings, and landscaping (Figure 12-6). The remotely sensed data will most likely capture the terrain in all 10 stages of development. Many of these stages may appear spectrally similar to other phenomena. For example, it is possible that stage 10 pixels (subdivided, paved roads, building, and completely landscaped) may look exactly like stage 1 pixels (original land cover) in multispectral feature space if a relatively coarse spatial resolution sensor system such as the Landsat TM (30×30 m) is used. This can cause serious change detection problems. Therefore, the analyst must be intimately aware of the phenological cycle of all urban phenomena being investigated, as well as the natural ecosystems.

The dichotomous key described in Figure 12-6 was used to identify residential development in January 25, 2007, and February 16, 2011 (near-anniversary date), high spatial resolution digital aerial photography of Beaufort County, SC (Figure 12-7). Once again, the land was *completely cleared* during the initial stages of development. Note the range of stages of development in the 2007 imagery including parcels with buildings in the process of being framed and parcels with just a foundation concrete pad (Figure 12-7a). Almost all of the residences in 2011 are completely landscaped (Figure 12-7b). The 2011 imagery is overlaid with 2012 parcel and street information. Each parcel with an address is highlighted with a flag icon (Figure 12-7b).

Progressive Stages of Residential Development

CLEARING FACTOR
Area been cleared at al?
no → 1. Original land cover is present
yes → Completely cleared?

SUBDIVISION FACTOR (yes)
Area subdivided?
no → 2. Area is cleared
yes →
TRANSPORTATION FACTOR
Roads paved?
no → BUILDING FACTOR
Buildings present?
no → 6. Cleared, subdivided, dirt roads
yes → LANDSCAPE FACTOR
Landscaping present?
no → 7. Cleared, subdivided, dirt roads, buildings
yes → Landscaping complete?
no → 8. Subdivided, dirt roads, buildings, partially landscaped
yes → 9. Subdivided, dirt roads, buildings, landscaped
yes → BUILDING FACTOR
Buildings present?
no → 3. Cleared, subdivided, paved roads
yes → LANDSCAPE FACTOR
Landscaping present?
no → 4. Cleared, subdivided, paved roads, buildings
yes → Landscaping complete?
no → 5. Subdivided, paved roads, buildings, partially landscaped
yes → 10. Subdivided, paved roads, buildings, landscaped

SUBDIVISION FACTOR (no)
Area subdivided?
no → 11. Area is partially cleared
yes →
TRANSPORTATION FACTOR
Roads paved?
yes → BUILDING FACTOR
Buildings present?
no → 12. Partially cleared, subdivided, paved roads
yes → LANDSCAPE FACTOR
Landscaping present?
no → 13. Partially landscaped, subdivided, paved roads, buildings
yes → 17. Subdivided, paved roads, buildings, landscaped
no → BUILDING FACTOR
Buildings present?
no → 14. Partially cleared, subdivided, dirt roads
yes → LANDSCAPE FACTOR
Landscaping present?
no → 15. Partially landscaped, subdivided, dirt roads, buildings
yes → 16. Subdivided, dirt roads, buildings, landscaped

FIGURE 12–6 Dichotomous key used to identify progressive stages of residential development. Such development in Denver, CO, normally begins by clearing the terrain of vegetation prior to subdivision. In many geographic areas, however, some natural vegetation may be kept as landscaping. The absence or existence of natural vegetation dramatically affects the range of signatures that a parcel of land undergoes as it progresses from natural vegetation (1) to fully-landscaped residential housing (10 or 17).

Obscuration Considerations

Sometimes the features of interest in a change detection investigation such as buildings, roads, agricultural land, stands of trees, etc. are obscured from view on one or both of the images being analyzed by vegetation or intervening buildings. For example, Figure 12-8a is a nadir view of several single-family residences in Beaufort, SC, collected on February 2, 2011. The image is centered on a home that is almost completely obscured from view by thick live oak canopy. The oblique views of the home looking from the North to the South (Figure 12-8c) and from the East to the West (Figure 12-8e) also do not provide much information of value. Conversely, the oblique views from the South to the North (Figure 12-8b) and from the West to the East (Figure12-8d) provide sufficient image information to determine that there is a residential home beneath the tree canopy. Note that the trees in the neighborhood also obscure some of the roads. Only a few companies such as Pictometry International, Inc. have software that can be used to detect change in buildings and roads from oblique imagery obtained on multiple dates.

Effects of Tidal Stage on Change Detection

Tidal stage is a crucial factor when conducting change detection in the coastal zone (Klemas, 2011). Ideally, the tidal stage is identical on multiple-date images used for change detection. Sometimes this severe constraint

Stages of Residential Development near Beaufort, SC, from January 25, 2007, to February 16, 2011

a. Stages of residential development identified in Pictometry 2007 imagery based on the dichotomous key in Figure 12-6 overlaid with 2012 Beaufort County, SC, parcel information.



b. Stages of residential development identified in 2011 imagery based on the dichotomous key in Figure 12-6 overlaid with Beaufort County parcels and street names.

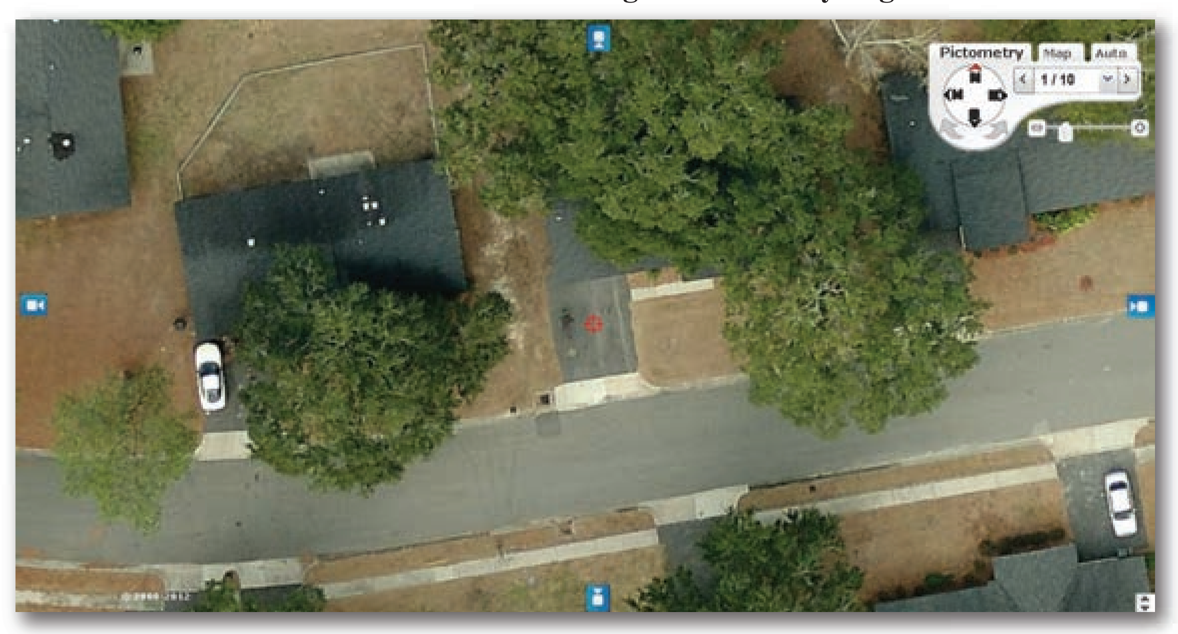
FIGURE 12–7 a) The stages of residential development described in Figure 12-6 applied to an area in Beaufort County, SC, in 2007. b) The stages of residential development for the same geographic area in 2011. Many of the parcels in 2011 are subdivided, have access to paved roads, buildings are present, and the property is fully landscaped (i.e., residential development Stage 10). Each building with an icon has a Beaufort County residential address (imagery courtesy of Beaufort County GIS Department).

can rule out the use of satellite remote sensing systems that cannot collect data off-nadir to meet the stringent tidal requirements. In such cases, the only way to obtain remote sensor data in the coastal zone that meets the stringent tidal requirements is to use suborbital sensors that can be flown at the exact time required. For most regions, images to be used for change detection acquired at mean low tide (MLT) are preferred, 1 or 2 ft. above MLT are acceptable, and 3 ft. or more above MLT are generally unacceptable (Jensen et al., 1993a; Klemas, 2011).

Select the Most Appropriate Change Detection Algorithm

The selection of an appropriate change detection algorithm is one of the most important considerations in the

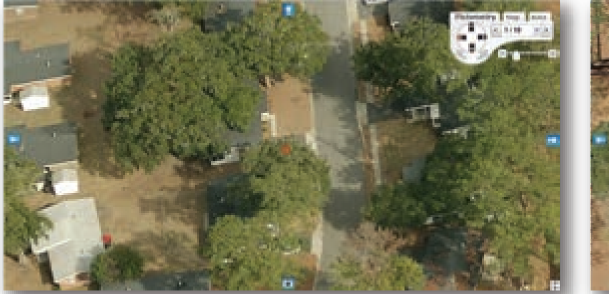
Obscuration of Residential Buildings and Roads by Vegetation



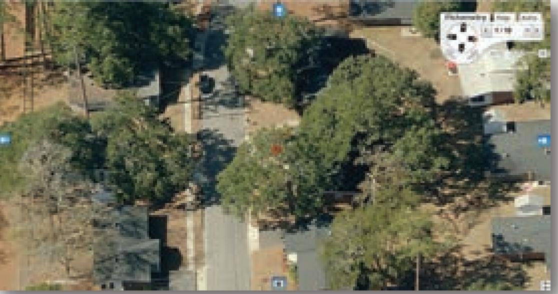
a. Image collected at Nadir.



b. Oblique image looking from the South to the North. c. Oblique image looking from the North to the South.



d. Oblique image looking from the West to the East.



e. Oblique image looking from the East to the West.

FIGURE 12–8 a) Buildings, roads and other terrain may be obscured from view by tall dense tree canopy when imagery is collected at Nadir in leaf-on conditions. In this example the image is centered on the driveway of an obscured single-family residence. b-e) The same residential area viewed from four different cardinal directions. The view from the South to the North (b) provides the most detailed information about the residential home (imagery courtesy of Beaufort County GIS Department).

change detection process. There is no single change detection algorithm that is suitable for all change detection projects (Chen et al., 2012). Digital change detection algorithms can provide binary land cover "change/no-change" information or more detailed thematic "from-to" change information which identifies changes in land cover from forest, agriculture, etc., into another type of land cover such as residential housing, apartment complexes, roads, etc. (e.g., Lu et al., 2004; Im et al., 2009, 2012; Tsai et al., 2011). The following sections describe several of the most important change detection algorithms.

Analog On-screen Visualization Change Detection

Original bands: blue, green, red, NIR (Date 1); blue, green, red, NIR (Date 2)

Transformed bands: $NDVI = \frac{red - NIR}{red + NIR}$ (Date 1); $NDVI = \frac{red - NIR}{red + NIR}$ (Date 2)

a. Date 1 red → Memory bank red; Date 2 red → Memory bank green, blue

b. NDVI Date 1 → Memory bank red; NDVI Date 2 → Memory bank green, blue

Advantages
* Visual examination of 2 or 3 years of nonspecific change
* Requires date-to-date histogram matching, not atmospheric correction

Disadvantages
* Nonquantitative
* No "from–to" change class information

FIGURE 12–9 The logic of analog visualization on a computer screen using two or three dates of registered remote sensor data placed in the RGB memory banks. Analysts can use a) individual original bands, and/or b) multiple-date transformations of the original data such as the Normalized Difference Vegetation Index (NDVI) (imagery courtesy of Richland County GIS Department).



Binary Change Detection Algorithms Provide "Change/No-Change" Information

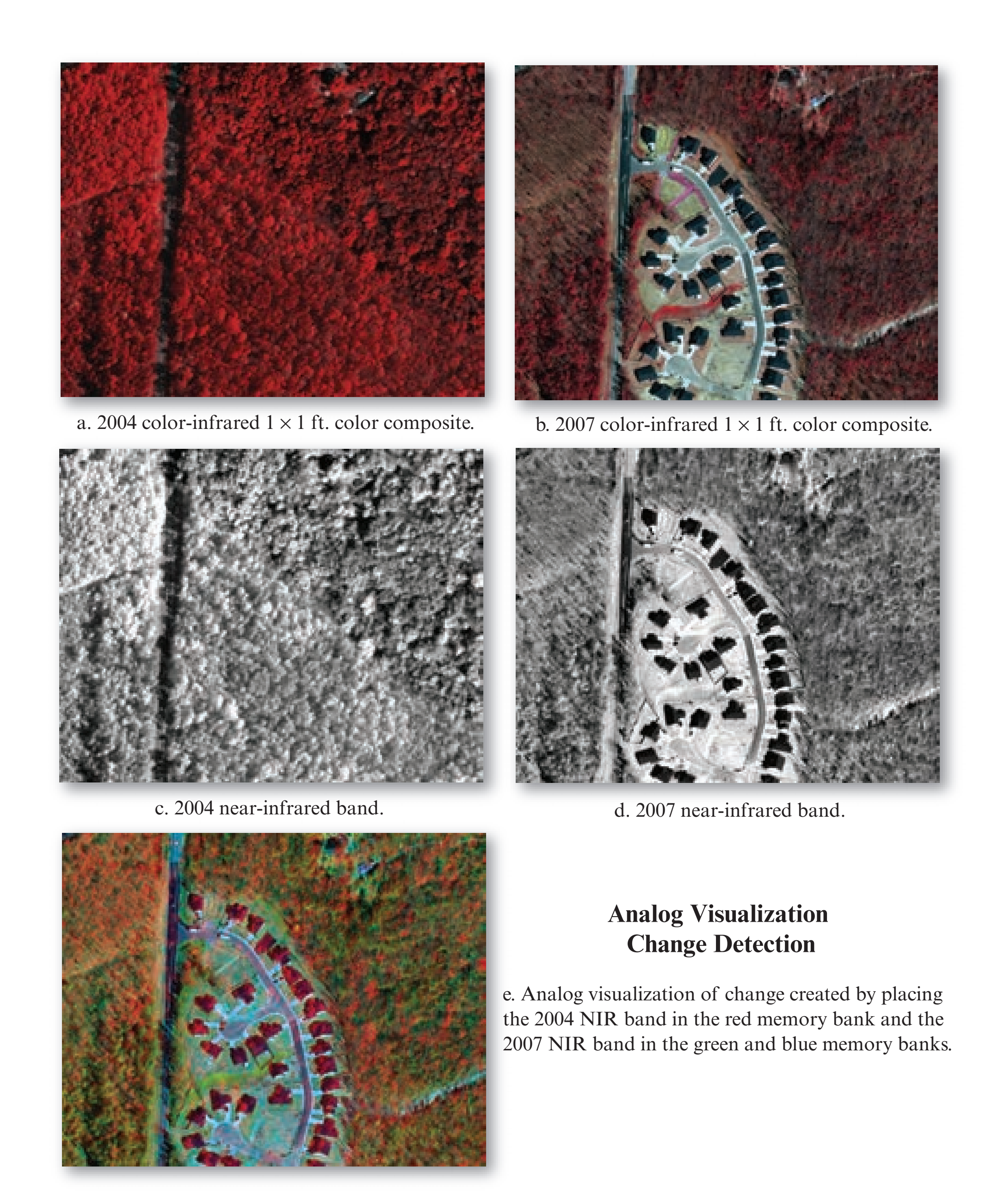
Two very useful binary change detection algorithms are: 1) analog "on-screen" visualization change detection (Figure 12-9), and 2) binary change detecting using image algebra (e.g., band differencing and ratioing) or Principal Components Analysis (PCA) composite image analysis.

Analog "On-Screen" Visualization Change Detection

Analog on-screen visualization change detection typically does not provide any quantitative change detection information. Rather, change information is displayed on the computer screen in various colors for visual interpretation and understanding.

Analog visualization involves the use of the three banks of graphics memory (RGB = red, green, and blue) located on the digital image processing system's computer graphics card (refer to Chapter 3 for a discussion about graphics memory). Basically, individual bands (or derivative products such as the Normalized Difference Vegetation Index—NDVI) from multiple dates of imagery are systematically inserted into each of the three memory banks (red, green, and/or blue) (Figure 12-9b) to highlight any changes in the imagery (Lunetta et al., 2006).

For example, consider the 2004 (Date 1) and 2007 (Date 2) digital aerial photography of an area in Ballentine, SC (Figure 12-10). The 2004 color-infrared color composite 1 × 1 ft. spatial resolution image exhibits complete canopy closure with little urban infrastructure present, except for a paved road and a few individual buildings (Figure 12-10a). The 2007 color-infrared image records a new residential subdivision (Figure 12-10b). The 2004 and 2007 near-infrared bands are shown in Figure 12-10c and d. A screen capture of the analog visualization of the change between the two dates is displayed in Figure 12-10e. It was created by placing the 2004 near-infrared band in the red



a. 2004 color-infrared 1 × 1 ft. color composite.

b. 2007 color-infrared 1 × 1 ft. color composite.

c. 2004 near-infrared band.

d. 2007 near-infrared band.

e. Analog visualization of change created by placing the 2004 NIR band in the red memory bank and the 2007 NIR band in the green and blue memory banks.

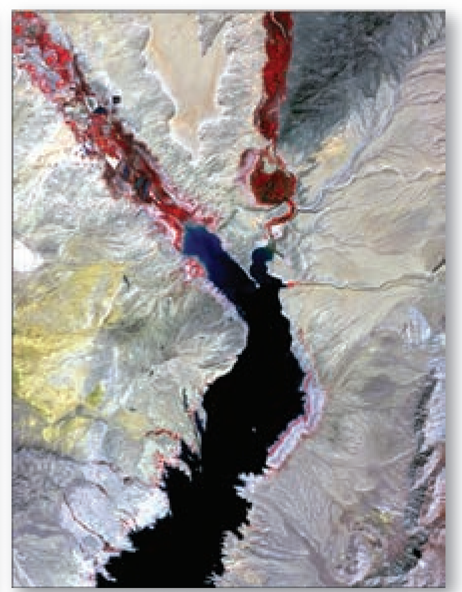
FIGURE 12–10 The use of analog visualization to detect changes in high spatial resolution imagery collected in 2004 and 2007 of an area near Ballentine, SC (imagery courtesy of Richland County GIS Department).

image processing video memory bank and the 2007 near-infrared band in both the green and blue image processing video memory banks. The result is a striking display of the residential change that has taken place in this area.

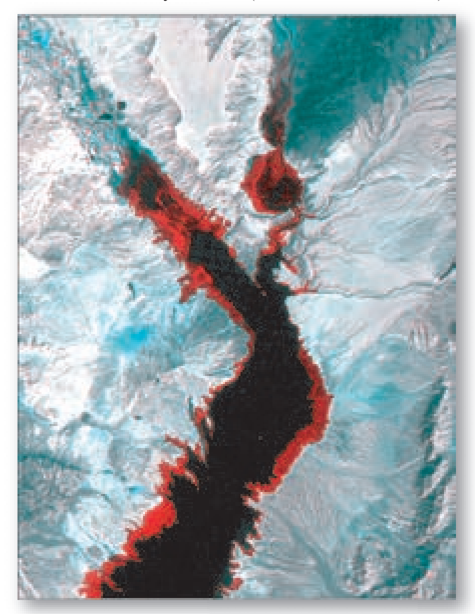
Analog visualization may be used to visually examine virtually any type of registered, multiple-date imagery. For example, Figure 12-11 demonstrates how two different types of multispectral data are analyzed to detect change. The Lake Mead, NV, watershed experienced severe drought from 2000 to 2003 resulting in a significant drawdown of the lake. Figure 12-11a depicts a color composite of Landsat ETM$^+$ imagery collected on May 3, 2000, and Figure 12-11b is a color composite of ASTER imagery collected on April 19, 2003. Both datasets record approximately the same near-infrared radiant flux and the images are within 14 days of being anniversary dates. The ETM$^+$ and ASTER images were resampled to 30 × 30 m (nearest-neighbor; RMSE $\pm$ 0.5 pixel). Analog visualization was performed by placing the ASTER band 3 image in the red memory bank, the ETM$^+$ band 4 data in the green memory bank, and the ETM$^+$ band 4 data in the green memory bank. This results in a clear depiction of the land exposed due to lake drawdown in shades of bright red.



a. Landsat ETM$^+$ data of Lake Mead, NV, obtained on May 3, 2000 (RGB = bands 4, 3, 2).



b. ASTER data obtained on April 19, 2003 (RGB = bands 3, 2, 1).



c. Analog visualization change detection where RGB = ASTER band 3, ETM$^+$ band 4, and ETM$^+$ band 4, respectively.

Analog Visualization Change Detection of Lake Mead, NV, using Landsat ETM$^+$ and ASTER Imagery

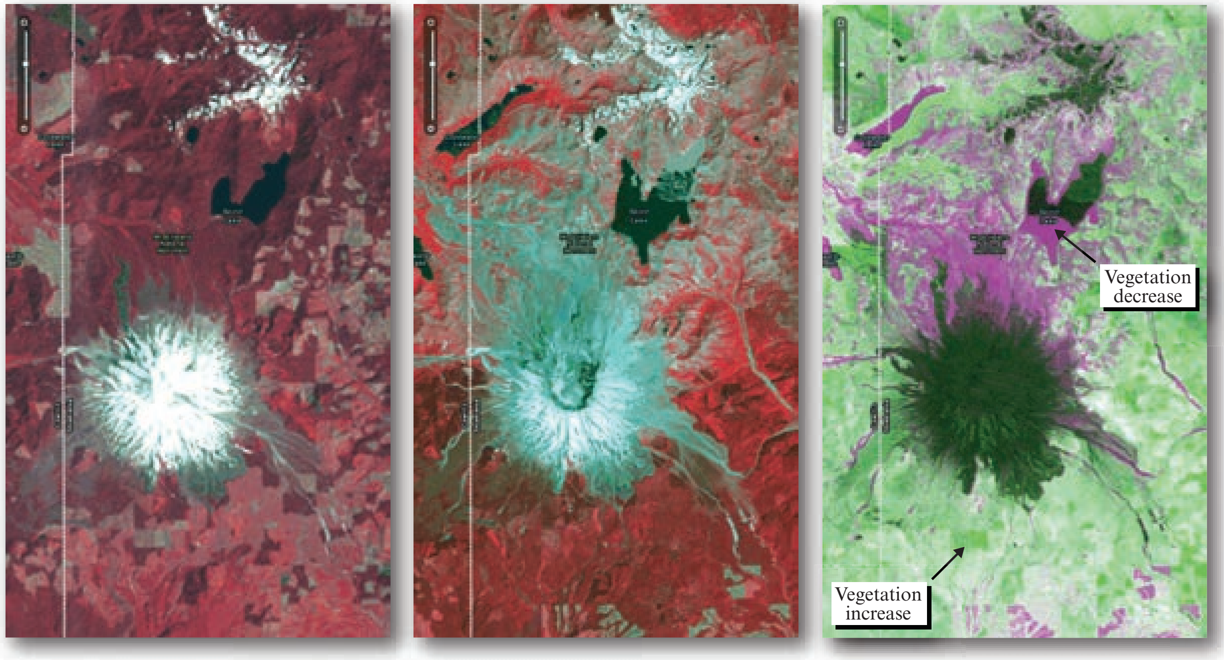
FIGURE 12–11 a,b) Multiple-date images of Lake Mead, NV, collected by Landsat ETM$^+$ and ASTER. c) Analog visualization change detection using ASTER band 3 (0.76 – 0.86 μm) and ETM$^+$ band 4 (0.75 – 0.90 μm) (imagery courtesy of NASA).

Esri, Inc., ChangeMatters®

Esri's ChangeMatters® Internet change detection program builds upon analog visualization logic to display change associated with multiple dates of remote sensor data, especially Landsat Multispectral Scanner (MSS), Thematic Mapper (TM), and ETM$^+$ imagery. ChangeMatters® operates as a free Landsat viewer web service that is built on ArcGIS Server Image extension (Esri ChangeMatters, 2014). It calculates the change on the fly rather than storing the results that have been preprocessed.

In December, 2008, the USGS made the entire Landsat archive, over 3 million images, available via the Internet at no cost (USGS, 2014). The USGS and NASA partnered to post the Global Land Survey (GLS) (http://landsat.usgs.gov/science_GLS.php) epochs of Landsat imagery, comprised of one image per each epoch worldwide from the 1970s, 1990s, 2000, 2005, and 2010.

Esri ChangeMatters® Analysis of Mt. St. Helens from 1975 to 2010



a. 1975 Landsat MSS image. b. 2010 Landsat TM image. c. NDVI Change 1975 to 2010.

FIGURE 12–12 Output from the Esri, Inc., ChangeMattersfi used to identify change on Mt. St. Helens from 1975 to 2010. a) A 1975 Landsat Multispectral Scanner (MSS) image (RGB = bands 4,2,1). b) A 2010 Landsat Thematic Mapper image (RGB = bands 4, 3, 2). c) A 2010 NDVI image was placed in the green graphics memory and a 1975 NDVI image was placed in the red and blue graphics memory. Unchanged areas appear in shades of grey because the earlier and later images should have approximately the same NDVI values.The NDVI change image highlights areas with vegetation decrease in magenta and vegetation increase in light green (images courtesy of USGS and NASA).

The ChangeMatters® program dynamically mosaics images in different projections, formats, locations, pixel sizes, and time periods (Green, 2011). The server-based processing enables on-the-fly creation of multiple image products with no intermediate files being generated. The application enables users to view different band combinations, different epochs of imagery, and to process multiple bands against one another to derive new products. Users can define what processing is to be performed on the imagery, and the server performs this directly on the source images, returning the information required for the area of interest.

ChangeMatters® displays are available in different standard band combinations including: near-infrared color composites (e.g.,TM bands 4, 3, 2) useful for vegetation studies and urban analysis; natural color with atmospheric penetration (e.g., TM bands 7, 4, 2) best suited for analysis of vegetation and some urban studies; healthy vegetation; land/water boundary detection; and vegetation analysis providing information for agriculture and forest management. Since these services are multitemporal, users can turn back the clock and easily analyze how the landscape has changed in the region over the past 30 years (Green, 2011).

We will use multiple dates of Landsat imagery of Mt. St. Helens, WA, to demonstrate the utility of the ChangeMatters® change detection program. Mount St. Helens is an active stratovolcano located in Skamania County, WA. It is 154 km south of Seattle, WA, and 80 km northeast of Portland, OR. Mount St. Helens is most famous for its catastrophic eruption on May 18, 1980, at 8:32 a.m. PDT. It is the deadliest and most economically destructive volcanic event in the history of the U.S. Fifty-seven people were killed; 250 homes, 47 bridges, 24 km of railways, and 298 km of highway were destroyed. A massive debris avalanche triggered by an earthquake measuring 5.1 on the Richter scale, caused an eruption, significantly reducing the elevation of the mountain's summit from 9,677 ft. (2,950 m) to 8,365 ft. (2,550 m) and replacing it with a 1 mile (1.6 km) wide horseshoe-shaped crater. The debris avalanche was up to 0.7 cubic miles (2.9 km^3) in volume. The Mount St. Helens National Volcanic Monument was created to preserve the volcano and allow for its aftermath to be scientifically studied.

A 1975 Landsat Multispectral Scanner (MSS) image with an original spatial resolution of 79 × 79 m was selected to be the pre-eruption image (Figure 12-12a). The post-eruption image was a 30 × 30 m Landsat TM

image collected in 2010 (Figure 12-12b). The change image (Figure 12-12c) is a composite of NDVIs extracted from these two dates of imagery. The 2010 NDVI image was placed in the green graphics memory bank and the 1975 NDVI image was placed in the red and blue graphics memory banks (Figure 12-12c) (Green, 2011). Unchanged areas appear in shades of grey because the earlier and later images should have approximately the same NDVI values. Areas with higher NDVI in the later year than the earlier year are green indicating an increase in vegetation vigor or a decrease in water level. Areas with lower NDVI in the later year than the earlier year are magenta (red + blue = magenta) indicating a decrease in vegetation vigor (perhaps to urban land cover) or an increase in water level. Access to ChangeMatters® on the Internet is free at http://changematters.esri.com/explore?tourid=10.

Another ChangeMatters example is presented as the introductory image to this chapter. It consists of 1990 and 2000 Landsat TM color-infrared color composites of the North Rim of the Grand Canyon. NDVI images were derived from these two TM datasets. The 2000 NDVI image was placed in the green graphics memory bank, the 1990 NDVI image was placed in the red and blue graphics memory banks. The analog visualization highlights the forest fire burn scar present in the 2000 Landsat TM image.

Advantages of analog visualization change detection include the possibility of looking at two and even three dates of remotely sensed imagery (or derivative products) at one time. Also, it is generally not necessary to atmospherically correct the remote sensor data used in analog visualization change detection. Unfortunately, the technique does not provide quantitative information about the amount of hectares changing *from* one land-cover category *to* another. Nevertheless, it is a useful analog method for qualitatively assessing the amount of change in a region, which might help with the selection of one of the quantitative change detection techniques.

Binary Change Detection Using Image Algebra

The extraction of quantitative binary "change/no-change" information using two dates of imagery is one of the most heavily used change detection methods (Im et al., 2009, 2011; Klemas, 2011). Binary change between two rectified images is typically performed using raster image algebra such as band differencing or band ratioing logic (Bruzzone and Prieto, 2000; Jensen and Im, 2007; Kennedy et al., 2009) or using a multiple-date composite image dataset and Principal Components Analysis (PCA) (Exelis, 2013; Rokni et al., 2015).

Image Differencing Change Detection

This technique involves subtracting band k of the Date 2 image from band k of the Date 1 image where the two bands (k) ideally have the same spectral resolution (e.g., a near-infrared band = 700 – 900 nm) (Figure 12-13). If the two images (bands) have almost identical radiometric characteristics (i.e., the band k data have been normalized or atmospherically corrected), the subtraction ideally results in positive and negative values in areas where radiance has changed and zero values in areas of no change in a *new* change image (Figure 12-13a).

The change image has a histogram with unique characteristics shown in Figure 12-13b (Im et al., 2011). When 8-bit data are analyzed, the potential range (domain) of difference values found in the change image is –255 to +255 (Figure 12-13b and 12-14). The results can be transformed into positive values by adding a constant, c (e.g., 127). The operation is expressed as:

$$\Delta BV_{i,j,k} = BV_{i,j,k}(1) - BV_{i,j,k}(2) + c \qquad (12.1)$$

where

$\Delta BV_{i,j,k}$ = change pixel value
$BV_{i,j,k}(1)$ = brightness value on date 1
$BV_{i,j,k}(2)$ = brightness value on date 2
c = a constant (e.g., 127)
i = line number
j = column number
k = a single band (e.g., IKONOS band 3).

The change image produced using image differencing usually yields a BV distribution approximately Gaussian, where pixels of no BV change are distributed around the mean and pixels of change are found in the tails of the distribution (Song et al., 2001). It is not necessary to add the constant c in Equation 12-1 if the image differencing output file is allowed to be floating point, i.e., the differenced pixel values can range from –255 to +255 (Figure 12-14a). Band ratioing involves the same logic, except a ratio is computed with values ranging from $\frac{1}{255}$ to 255 and the pixels that did not change have a ratio value of 1 in the change image (Figure 12-14a).

Image differencing change detection will be demonstrated using two datasets. The first example involves the Landsat ETM$^+$ imagery (band 4; 0.75 – 0.90 μm) obtained on May 3, 2000, and the ASTER imagery (band 3; 0.76 – 0.86 μm) of Lake Mead obtained on April 19, 2003 (Figure 12-15ab). These two images were differenced, resulting in the change image histogram shown in Figure 12-15c. Note that the change image histogram is symmetrical, suggesting that one image was histogram-matched (i.e., normalized) to the

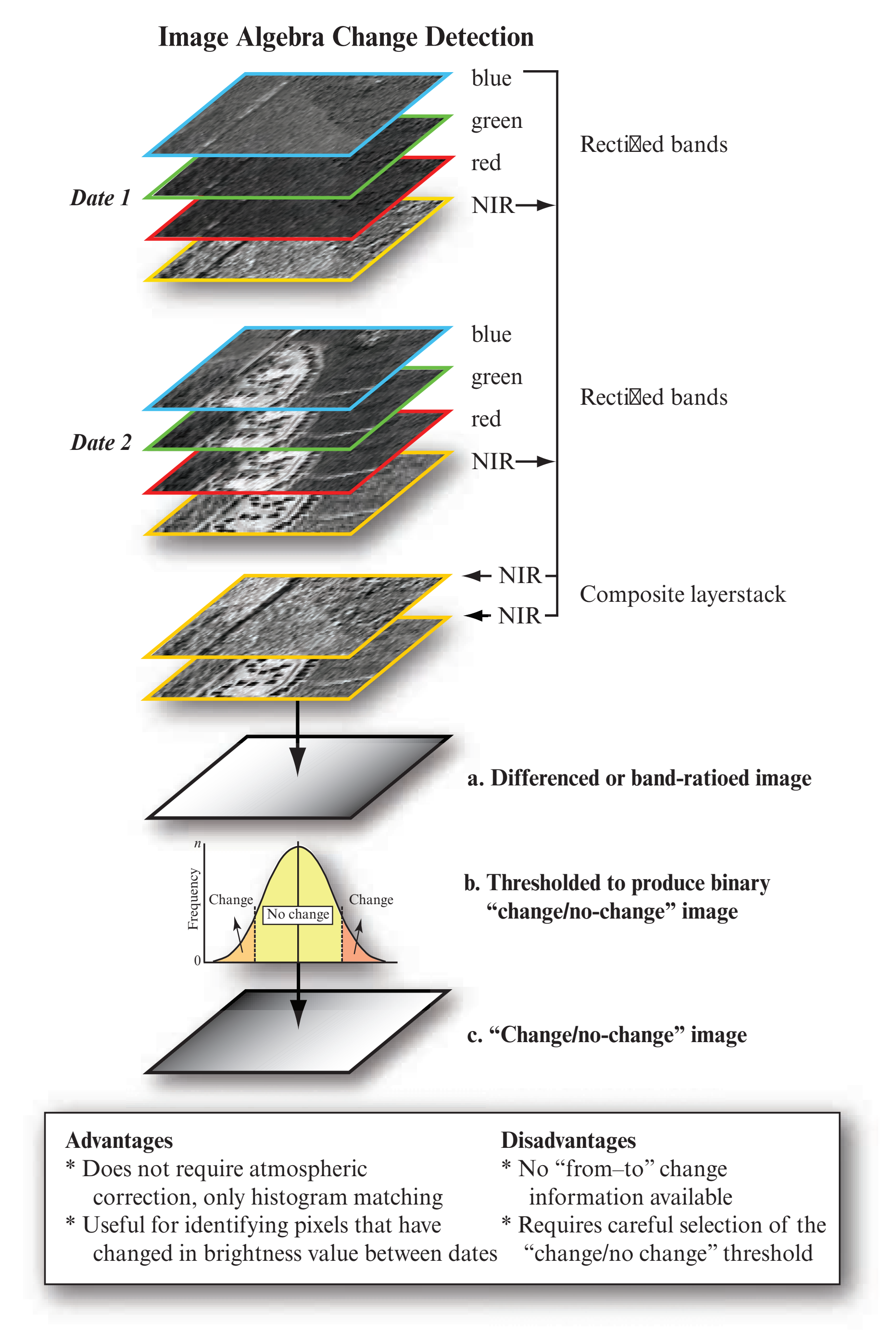


FIGURE 12–13 The logic of image algebra change detection using image differencing or band ratioing (imagery courtesy of Richland County GIS Department).

other (Kennedy et al., 2009). Most of the geographic area did not change between 2000 and 2003. Therefore, the vast majority of the pixels are found around the value 0 in the histogram. However, where the lake was drawn down, exposing bedrock, and where new vegetation has grown on the exposed terrain has resulted in significant change documented in the change image in Figure 12-15d. There are actually two types of change in Figure 12-15d, bright green and red. All pixels in the change histogram below the first analyst-specified threshold were assigned red and all those above the second threshold were assigned green.

Image Algebra Band Ratioing Change Detection

Binary change detection using band ratioing involves dividing band k of the Date 1 image by band k of the Date 2 image where band k ideally has the same spectral resolution:

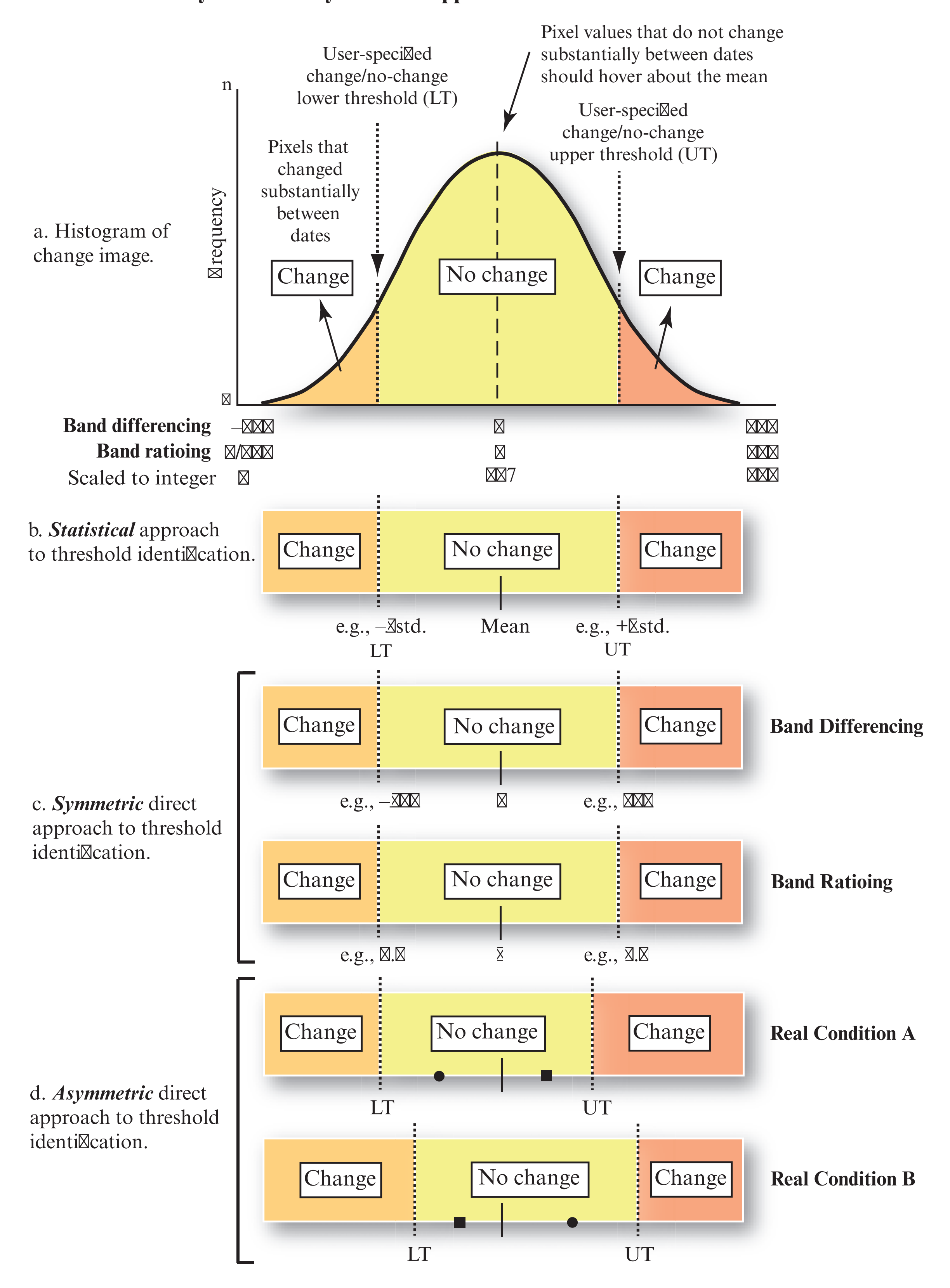
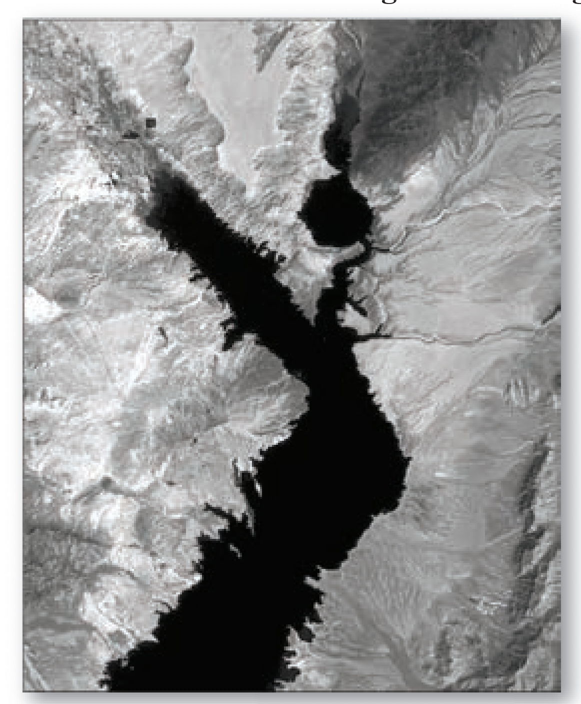
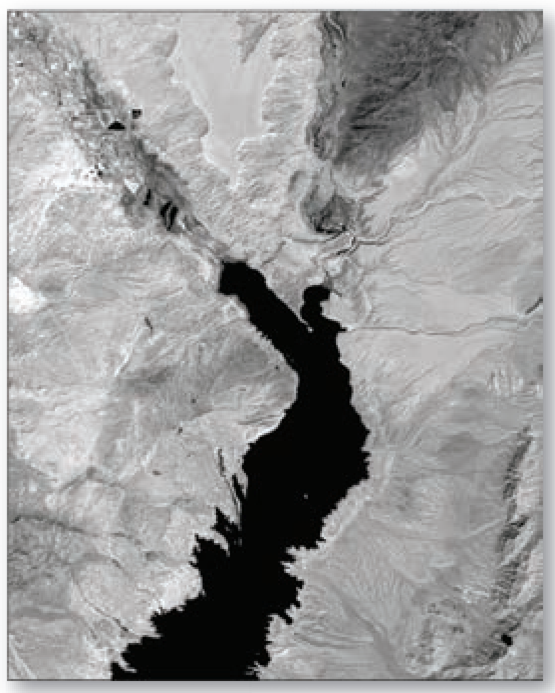


FIGURE 12–14 A change image created from multiple date imagery using band differencing or band ratioing can be processed to identify change using b) statistical, c) symmetric, or d) asymmetric approaches (based on Figure 1 from Im, J., Lu, Z., and J. R. Jensen, 2011, "A Genetic Algorithm Approach to Moving Threshold Optimization for Binary Change Detection," *Photogrammetric Engineering & Remote Sensing*, 77(2):167–180; and, Jensen et al., 2012).

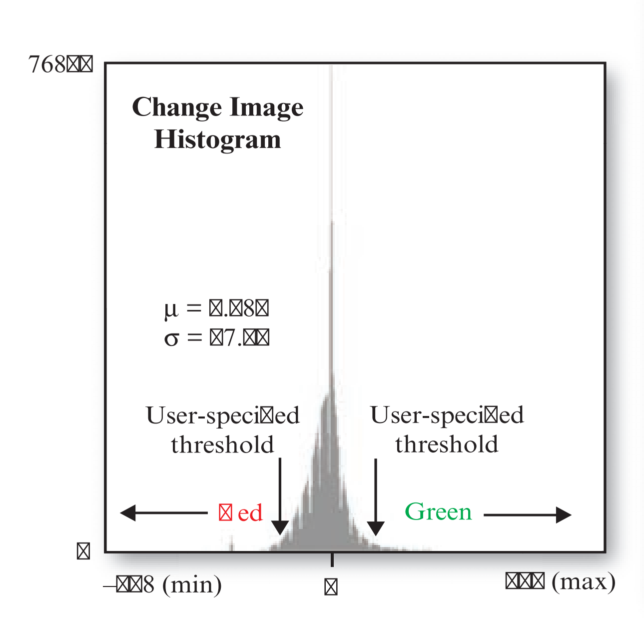
Image Differencing Change Detection



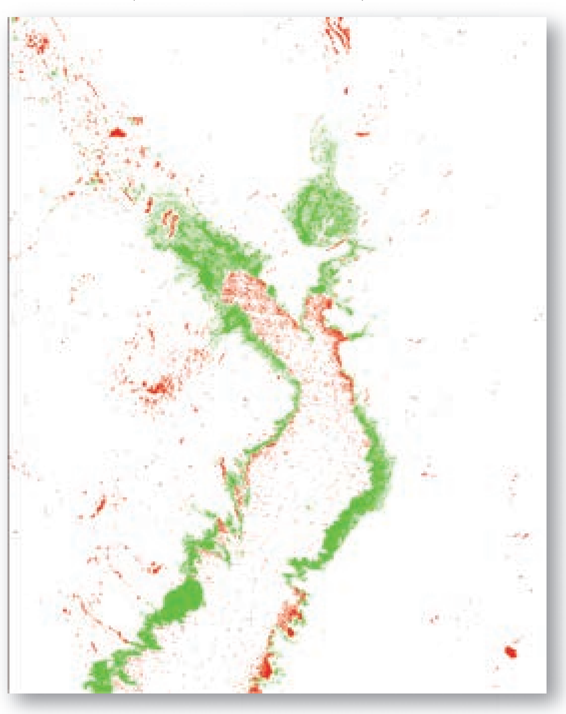
a. Landsat ETM$^+$ data of Lake Mead, NV, obtained on May ⊠, ⊠⊠⊠⊠ (⊠GB = bands ⊠, ⊠, ⊠).



b. ASTE⊠ data obtained on April ⊠⊠, ⊠⊠⊠⊠ (⊠GB = bands ⊠, ⊠, ⊠).



c. Histogram of the ⊠oating point image created by subtracting ⊠⊠⊠⊠ ETM$^+$ data from ⊠⊠⊠⊠ ASTE⊠ data. The symmetrical distribution con⊠rms that the images were histogram-matched prior to processing.



d. Image differencing change detection. ⊠ed and green change pixels are associated with thresholds located in the tails of the histogram.

FIGURE 12–15 a) Landsat ETM$^+$ imagery of a portion of Lake Mead in Nevada obtained on May 3, 2000. b) ASTER data of Lake Mead obtained on April 19, 2003. c) The histogram of a change image produced by subtracting the ETM$^+$ 2000 band 4 (0.75 – 0.90 ⊠m) data from the ASTER 2003 band 3 (0.76 – 0.86 ⊠m). d) Map showing the change as a function of the two thresholds identified in the change image histogram (original imagery courtesy of NASA Earth Observatory).

$$\Delta BV_{i,j,k} = \frac{BV_{i,j,k}(1)}{BV_{i,j,k}(2)} \quad (12.2)$$

Crippen (1988) recommended that all data used for multiple-date band ratioing be atmospherically corrected and free from any sensor calibration problems (e.g., a detector out of adjustment). If the two bands have almost identical radiometric characteristics, the resultant change image contains values that have a range (domain) from 1/255 to +255 (Figure 12-14a). Pixels that experienced minimal spectral change in band k on the two dates will have values that hover about the value 1.0. Unfortunately, the computation is not always simple because $BV_{i,j,k} = 0$ is possible in either date of imagery. Therefore, any $BV_{i,j,k}$ with a value of 0 is typically made equal to 1. Alternatively, a small value (e.g., 0.1) can be added to the denominator if it equals zero.

To represent the range (domain) of the change image data in a standard 8-bit format (values from 0 to 255), normalizing functions are usually applied. Using these normalizing functions, the ratio value 1 is assigned the brightness value 127. Change image ratio values within the range 1/255 to 1 are assigned values from 1 to 127 using the function (Jensen et al., 2012):

$$\Delta BV_{i,j,k} = Int[(BV_{i,j,r} \times 126) + 1] \quad (12.3)$$

Change image ratio values from >1 to 255 are assigned values within the range 128 to 255 using the function:

$$\Delta BV_{i,j,k} = Int\left(128 + \frac{BV_{i,j,r}}{2}\right) \quad (12.4)$$

Values in the ratio change image near 127 typically exhibit minimal change, while pixel values in the tails of the distribution have usually changed significantly between dates (Figure 12-14a).

It is important to note that band ratioing or image differencing change detection does not have to be based on just the individual bands of remote sensor data. It may also be extended to comparing vegetation or other type of index information derived from multiple dates of imagery. For example, scientists often compute an NDVI on two dates and then subtract one from another to determine change (e.g., Lunetta et al., 2006; Green, 2011; Klemas, 2011):

$$\Delta \mathrm{NDVI}_{i,j} = \mathrm{NDVI}_{i,j}(1) - \mathrm{NDVI}_{i,j}(2) + c \quad (12.5)$$

It is not necessary to use the constant c in Equation 12.5 if the image differencing output file is allowed to be floating point. The individual images used to accomplish NDVI change detection should ideally be atmospherically corrected. Change detection based on differencing multiple-date Kauth-Thomas transformations (e.g., change in brightness, greenness, and/or wetness) have also been widely adopted (e.g., Franklin et al., 2002; Radke et al., 2005).

Image Algebra Change Detection Using Statistical or Symmetric Thresholds

An important consideration when performing either image differencing or band ratioing binary change detection is deciding where to place the lower threshold (LT) and upper threshold (UT) boundaries between "change" and "no-change" pixels associated with the change image histogram (Im et al., 2011; Patra et al., 2013). The ideal threshold boundaries are rarely known *a priori*, but have to be determined. Sometimes a statistical approach using a standard deviation from the mean is selected and tested such as ±1 standard deviation (Figure 12-14b) (Morisette and Khorram, 2000). The statistical approach works well if change has occurred between dates in limited locations, e.g., ≤5% of the entire study site. If a larger portion of the study area has changed, the approach does not work well because the distribution of the pixel values is usually skewed. It is also possible to directly place the thresholds a symmetric distance from the midpoint as shown in Figure 12-14c.

Image Algebra Change Detection Using Asymmetric Thresholds

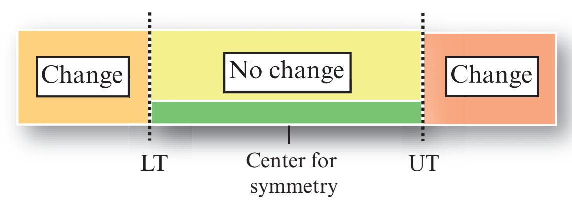
The direct symmetric approach generally works well if the radiometric correction of the remotely sensed data is near-perfect. However, near-perfect radiometric correction is rarely achieved due to factors described earlier in this chapter. Consequently, as illustrated in Figure 12-14d, asymmetric thresholds may work better than the symmetric thresholds for binary change detection using image differencing or band ratioing. Analysts often prefer to experiment empirically, placing the thresholds at non-symmetric locations in the tails of the change image histogram distribution until a realistic amount of change is encountered (Figure 12-14d). Thus, the amount of change selected and eventually recoded for display is often subjective and must be based on familiarity with the study area.

Image Algebra Change Detection Using Moving Threshold Windows (MTW)

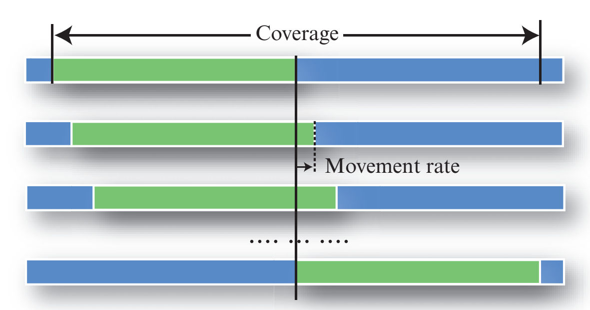
Research on binary change detection has also focused on the development of automated methods of identifying the optimum lower and upper thresholds when creating change images. For example, Im et al. (2009, 2011) developed moving threshold windows (MTW) to identify the optimum location of the thresholds. The logic associated with the MTW-based approach to binary change detection is illustrated in Figure 12-16. The approach consists of four basic processes. First it

Binary Change Detection Using A Moving Threshold Window (MTW) Approach

☒. Determine the initial size of the MT☒ .



☒. Calibrate with the initial size of the MT☒ .



☒. Calibrate with different sizes of MT☒ .

☒. Identify the optimum thresholds and size of the MT☒ resulting in the highest Kappa Coef☒cient of Agreement.

FIGURE 12–16 The logic associated with using a moving threshold window to identify the optimum thresholds during binary change detection (based on Figure 1 in Im et al., 2011, cited in Figure 12-14).

is necessary to identify the initial size of the MTW. The analyst can specify the initial size or accept a default initial size that is set as the nearest value to the center for symmetry in the domain of a difference-typed change-enhanced image. The default initial size for a ratio-typed image is set as the farthest value from the center for symmetry in the domain. The next step involves locating the threshold window in the domain and evaluating the numerous locations of the window using an automated calibration model (Figure 12-16). Two parameters (i.e., coverage and movement rate of MTW) are determined by the analyst. The coverage represents where a threshold window can move and the movement rate is the step size with which the window moves in the coverage. Since only one (initial) size of threshold window is calibrated in the second process, different (i.e., increasing or decreasing) sizes of MTWs need to be evaluated to see if they produce better binary change detection calibration. This is performed in the third process. The analyst can determine the maximum (or minimum) size of MTW. The maximum (or minimum) value of the domain of the corresponding change-enhanced image is usually set as the default (Figure 12-16).

Finally, the process identifies the optimum thresholds and size of MTW that result in the highest accuracy (Figure 12-16). The MTW-based model calculates three accuracies including user's and producer's accuracies for change, and the Kappa Coefficient of Agreement (Congalton and Green, 2009; Im et al., 2007, 2009). Patra et al. (2013) provide additional approaches to threshold selection when performing image differencing change detection.

Multiple-Date Composite Image Change Detection

Numerous researchers have rectified multiple dates of remotely sensed imagery (e.g., selected bands of two IKONOS scenes of the same region) and placed them in a single *layerstack* (Figure 12-17). This composite dataset can then be analyzed in a number of ways to extract change information (e.g., Deng et al., 2008; Kennedy et al., 2009; Rokni et al., 2015).

Supervised and Unsupervised Classification of the Multiple-Date Composite Image to Detect Change

First, a traditional classification using all n bands (eight in the example in Figure 12-17b) may be performed. Unsupervised classification techniques will result in the creation of numerous clusters. Some of the clusters contain change information while others do not. The analyst must then use careful judgment to label the clusters accordingly.

Principal Components Analysis (PCA) Composite Image Change Detection

PCA can be used to detect change in multiple dates of remote sensor data by merging the Date 1 and Date 2 image datasets into a single multi-temporal composite image *layerstack* (Figure 12-17a). The data stored in the multispectral layerstack are usually highly correlated with one another. The PCA evaluates the variance-covariance matrix associated with the n bands in the layerstack (yielding an un-standardized PCA) or the correlation matrix associated with the n bands in the layerstack (yielding a standardized PCA). The PCA determines a new set of orthogonal axes that have their origin at the data mean and that are rotated so the data variance is maximized (Nielsen and Canty, 2008; Celik, 2009; Kennedy et al., 2009; Almutairi and Warner, 2010; Exelis, 2013). The output from a typical PCA is an uncorrelated set of n component images that account for increasingly smaller amounts of variance (Figure 12-17c).

The output component images can be analyzed using two different methods. First, a traditional classification using all n component images (six in the example in Figure 12-17d) may be performed. Unsupervised, supervised, or hybrid classification techniques are used to

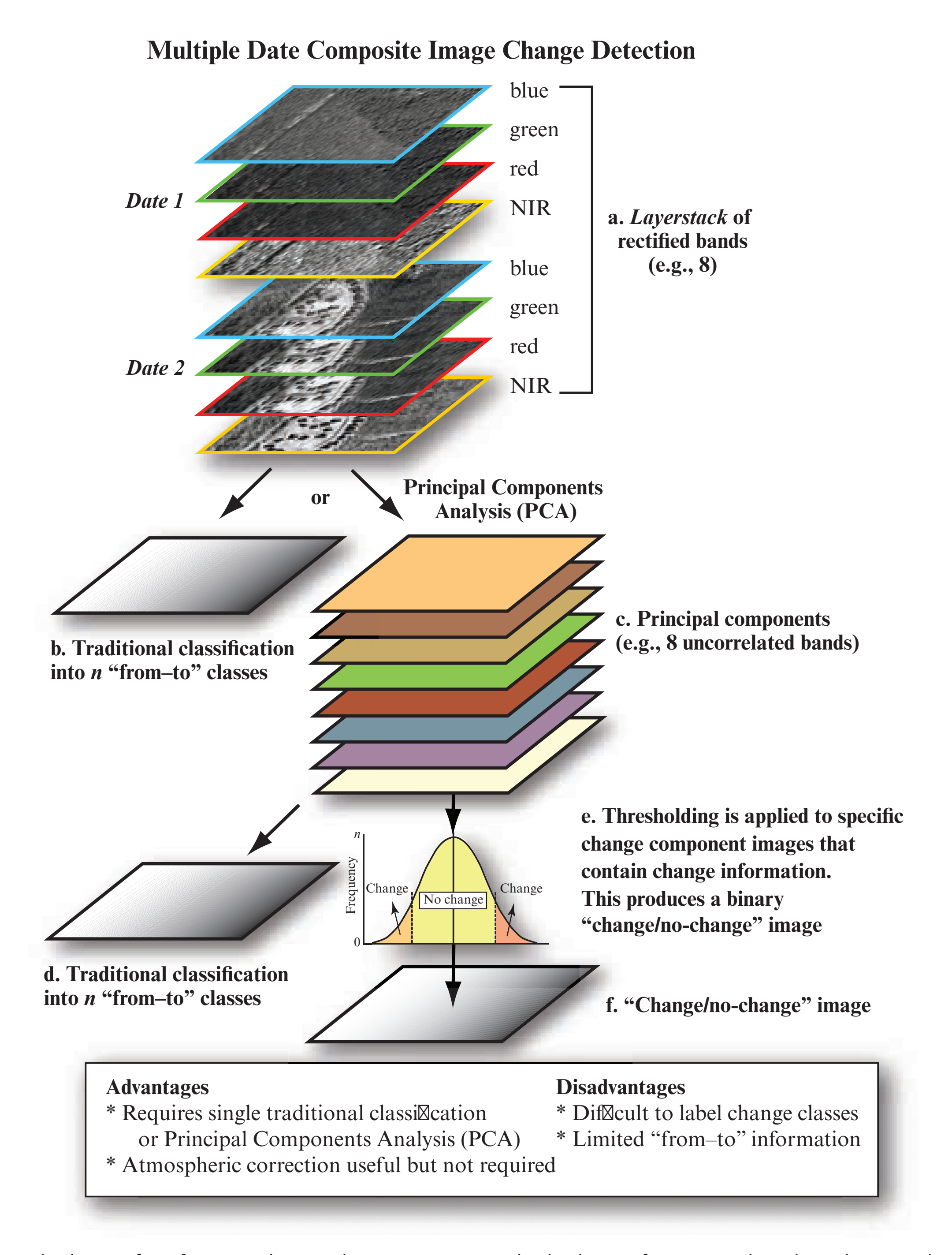


FIGURE 12–17 The logic of performing change detection using multiple dates of imagery placed in a layerstack. The layerstack can then be analyzed using b) traditional unsupervised or supervised classification techniques, or c) by using Principal Components Analysis (PCA) (imagery courtesy of Richland County GIS Department).

identify clusters in feature space associated with change (e.g., Rokni et al., 2015). The analyst must then label the clusters, which can be difficult (Deng et al., 2008). This method does yield some "from–to" information if it can be deciphered.

A more common approach is to select and analyze just the PCA component (i.e., band) that contains most of the important change information (Figure 12-17c,e). The first one or two PCA images generated from the multiple-date layerstack tend to account for variation in the image data that is not due to land-cover change. These are called stable components. When all of the bands are used, change information is often found in the third or fourth PCA band, but it varies depending on multiple-date scene content and the amount of change present between the two images. PCA bands that contain change information are called *change components* (Collins and Woodcock, 1996). The difficulty arises when trying to identify the ideal change

Principal Components Derived from a Multiple-Date Composite Image Dataset

Lake Mead, NV, dataset consists of:

* Landsat ETM+ data obtained on May 3, 2000 (bands 2, 3, 4)

* ASTER data obtained on April 19, 2003 (bands 1, 2, 3)

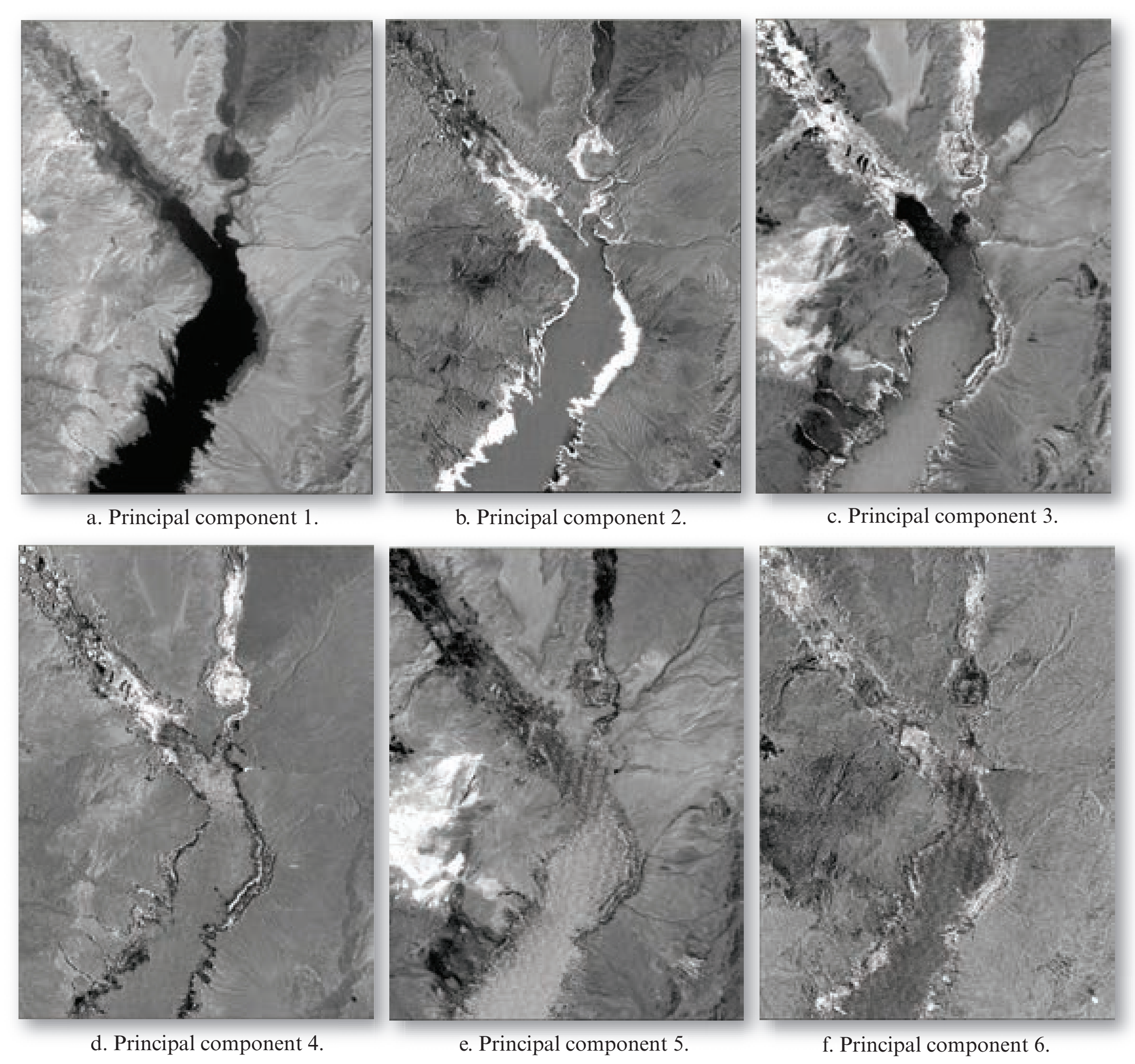


FIGURE 12–18 Principal components derived from a multiple-date dataset consisting of Landsat ETM^{+} and ASTER imagery. Principal component 2 contains change information. The first three principal components were placed in various memory banks to highlight more subtle changes, as shown in Figure 12-19 (original imagery courtesy of NASA).

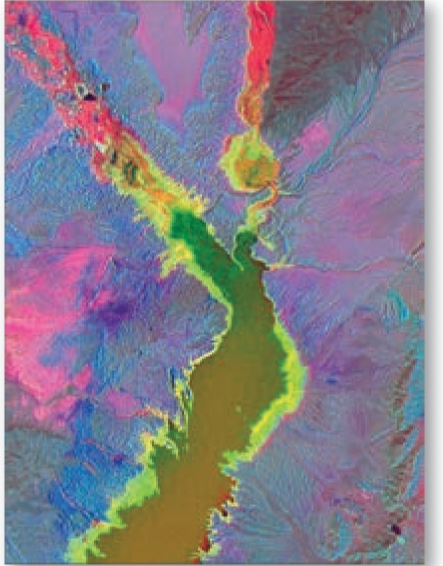
component image and extract change information (Almutairi and Warner, 2010). Once the PCA change component is identified, it is usually necessary to evaluate its histogram and identify the "change versus no-change" thresholds in the tails of the distribution to isolate the change as previously discussed (Figure 12-17e). Change areas are typically highlighted as bright and/or dark pixels (Figure 12-17f).

An example of a multiple-date composite image change detection will be demonstrated using two three-band Lake Mead, NV, datasets (Landsat ETM^{+} and ASTER data). The datasets were merged into a single six-band dataset and subjected to a principal components analysis. This resulted in the creation of the six principal component images shown in Figure 12-18. Note that principal component 2 is a *change component image* (Figure 12-18b) containing detailed information about the area exposed by the lake drawdown. The histogram of the principal component 2 image in Figure 12-18b could easily be analyzed and a change map created (not shown). More subtle change information can be visually extracted from the multiple-date component dataset by placing each of the first three principal components in various memory banks, as shown in Figure 12-19.

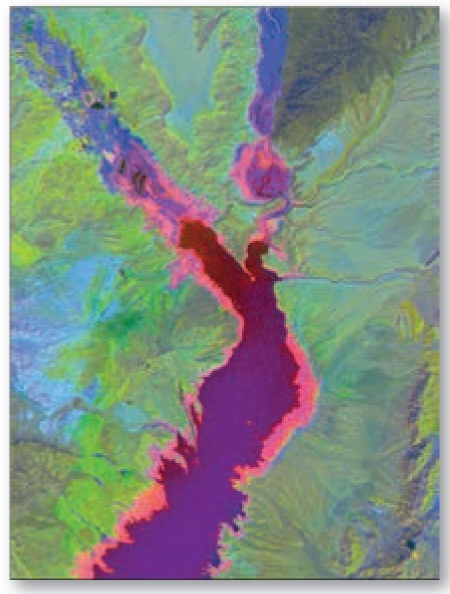
Multiple-Date Composite Image Change Detection Based on Principal Components Viewed Using Analog Visualization



a. Analog visualization change detection using RGB = principal components 3, 2, 1.



b. Analog visualization change detection using RGB = principal components 1, 2, 3.



c. Analog visualization change detection using RGB = principal components 2, 1, 3.

Principal Component Images Derived from a Multiple-Date Composite Dataset Consisting of

* Landsat ETM+ data obtained on May 3, 2000 (bands 2, 3, 4)

* ASTER data obtained on April 19, 2003 (bands 1, 2, 3)

FIGURE 12–19 Multiple-date composite image change detection using PCA images that were created from a multiple-date composite dataset consisting of Landsat ETM+ and ASTER data. Analog visualization was used to highlight the information content of the various principal component images. Principal components 2 and 3 contain much of the change information (original imagery courtesy of NASA).

MDA Information Systems LLC, National Urban Change Indicator (NUCI)fi

Another approach to the utilization of Landsat data for change detection is offered by MDA Information Systems LLC (MDA, 2014). Unlike the ChangeMatters application, NUCIfi utilizes both raster and vector GIS data that have been preprocessed using MDA's proprietary Correlated Land Change (CLC)fi processes (Dykstra, 2012). The NUCI process systematically compares the spectral characteristics of Landsat data acquired over the past years to determine when a specific 30 × 30 m pixel has experienced a "persistent" change. One of the output GIS data files includes the date when a change has occurred for each 30 × 30 m pixel in the entire contiguous 48 states.

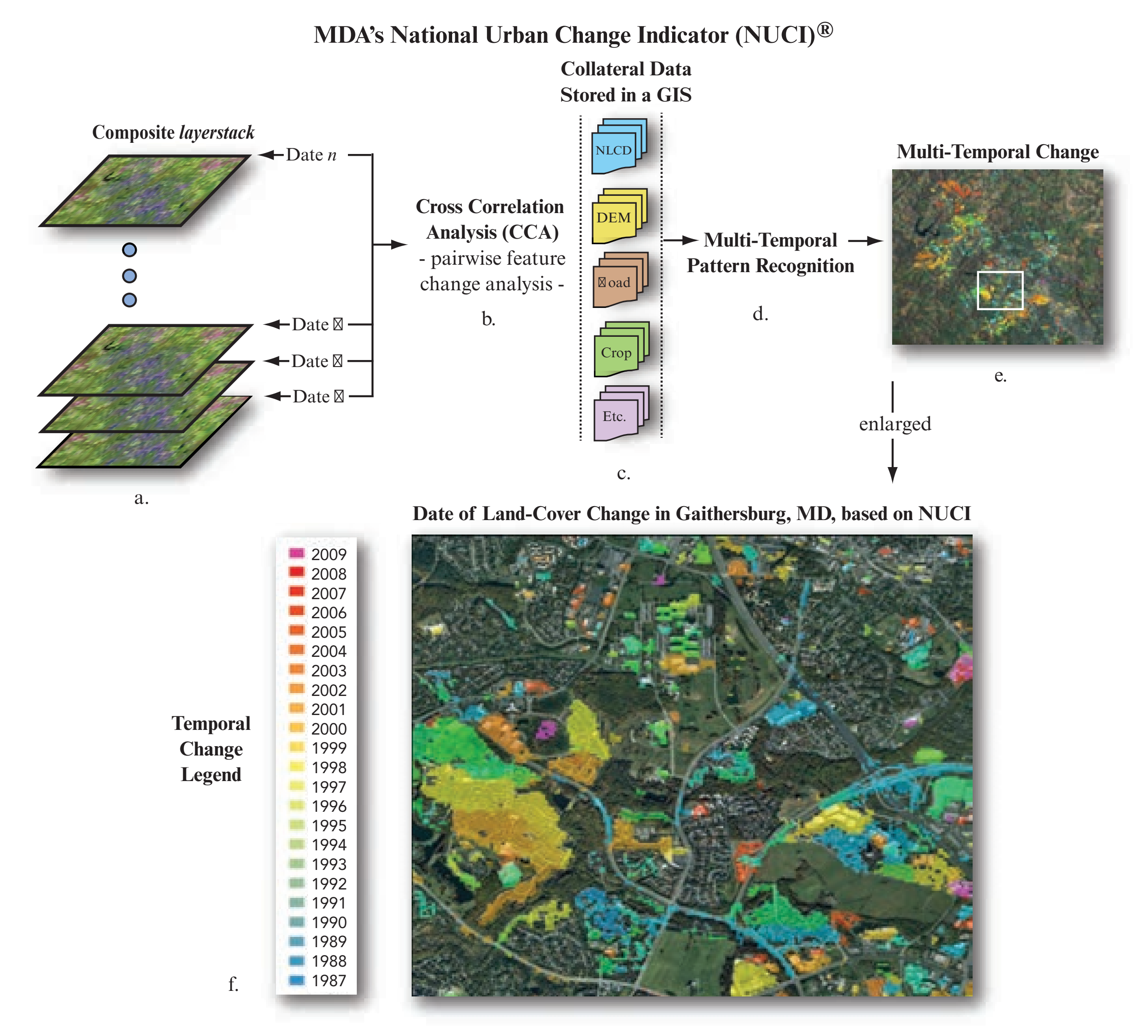


FIGURE 12–20 a) MDA Information Systems, LLC's National Urban Change Indicator (NUCI) program uses registered Landsat (or other) data. b) Cross Correlation Analysis (CCA) is performed using imagery, and c) collateral geospatial data. d) Pattern recognition using CCA-derived and collateral data. e) Color-coded date of land-cover change based on NUCI analysis of multi-temporal Landsat data of Gaithersburg, MD. f) Enlargement with temporal change legend from 1987 to 2009 (adapted from the internet document, Dykstra, J., 2012, "Comparison of Landsat and RapidEye Data for Change Monitoring Applications," *11th Annual Joint Agency Commercial Imagery Evaluation Workshop*, Fairfax, VA, April 17–19, 32 p).

The input data for the Correlated Land Change process is a multi-temporal *layerstack* of co-registered Landsat multispectral scenes as shown in Figure 12-20a. MDA uses proprietary Cross Correlation Analysis (CCA) to determine pixel by pixel change by analyzing two-date change detection results calculated for a systematic collection of pairwise permutations across a multi-temporal layerstack of precision co-registered Landsat multispectral data (Figure 12-20b). A multi-date pattern recognition process is then used to determine and map urban and other anthropomorphic change on a pixel-by-pixel basis (Figure 12-20c-e). NUCI change can be reported in two classes:

1. A *3-observation change* is associated with pixels that remain in a consistent state (no measured change) for three consecutive dates in the multi-temporal layerstack; then change to a different state and remain in that different state for another three consecutive dates. This is determined by observing that each of the three observations within the new state showed, for that pixel, change with all three observation of the earlier state, while all pixels within a given state do not show change among themselves.
2. A *2-observation change* (called potential recent change) dataset contains pixels that exhibit a two-date change within the multi-temporal layerstack's most recent five observations. A pixel is flagged as a potential recent change when two of the most recent observations of the multi-temporal image

TABLE 12–3 Change Attribute Majority (CAM) classes (adapted from Dykstra, J., 2012, cited in Figure 12-20).

1. Probable Urban Change
2. Probable Water Change
3. Probable River/Stream Change
4. Probable Shadow Change
5. Probable Crop Change
6. Change away from Roads
7. Vegetated Probable Urban Change
8. Vegetated Probable Water Change
9. Vegetated Probable River/Stream Change
10. Vegetated Probable Shadow Change
11. Vegetated Probable Crop Change
12. Vegetated Change away from Roads

layerstack show no change with each other, and each showed change with a group of three earlier internally consistent (showing no change among the three) observations. Such a pixel could become a 3-observation change if a future observation measured no change with the upper pair within the temporal stack while, at the same time, showed change with each of the same three older observations.

The Cross Correlation Analysis procedure uses an MDA-defined triply confirmed multi-temporal test to identify "persistent" change (Figure 12-20b). Most persistent change is associated with urban/cultural features. However, there are other phenomena that can trigger a persistent change (e.g., a change in reservoir level, recently cleared vegetation, strong seasonal shadows, etc.). Therefore, a collection of additional geospatial datasets is analyzed using GIS functions to "weed out" those pixels/polygons that are probably caused by non-urban phenomena (Figure 12-20c-e).

Probable Urban Change pixels are those that pass tests based on GIS overlay analysis, such as:

- Close to water but not close to NLCD-2006 defined urban, and
- Not in an area of high local relief, and
- Not coincident with USDA/NASS 2009 defined croplands, and
- Not a large distance away from TIGER_2010 S1400-level roads.

The NUCI change data are available both as a raster and polygon representation of single or contiguous cells that have undergone change. These changes are categories representing the most probable cause of the NUCI detected change. The Change Attribute Majority classes are summarized in Table 12-3.

The date and spatial distribution of land-cover change in Gaithersburg, MD, from 1987 to 2009 based on the National Urban Change Indicator (NUCI) analysis is shown in Figure 12-20f (Dykstra, 2012). Several agencies such as the Dept. of Defense, National Geospatial–Intelligence Agency, and the Dept. of Homeland Security make use of NUCI data. MDA Information Systems LLC will include Landsat 8 data into NUCI. Access to the NUCI database is by subscription.

Continuous Change Detection and Classification (CCDC) using Landsat Data

Remote sensing of the Earth has been conducted routinely from some satellites for more than four decades (e.g., Landsat 1-8). Scientists are developing algorithms that can mine these multi-temporal datasets to identify the change taking place based on the use of every image available, instead of just two images at a time. For example, Zhu and Woodcock (2014) developed a new algorithm for Continuous Change Detection and Classification (CCDC) of land-cover based on the use of *all* available Landsat data. The algorithm is capable of detecting many kinds of land cover change continuously as new images are collected and providing land-cover maps for any given time. Cloud and snow masking algorithms are used to eliminate "noisy" observations. The CCDC algorithm uses a threshold derived from all seven Landsat bands. When the difference between observed and predicted images in a multi-temporal dataset exceeds a threshold three consecutive times, a pixel is identified as land surface change. The actual land-cover classification is performed after change detection using a Random Forest Classifier.

Zhu and Woodcock (2014) found that the CCDC results applied to Landsat data of New England were accurate for detecting land surface change, with a producer's accuracy of 98% and a user's accuracy of 86% in the spatial domain and a temporal accuracy of 80%. The CCDC methodology is computationally intensive. Also, the algorithm requires a high temporal frequency of clear observations. For places with persistent snow or clouds, the CCDC model estimates may not be accurate as there are not enough observations to estimate the time series model.

Similarly, Huang and Friedl (2014) developed a land-cover change detection algorithm that utilizes more than 10 years of MODIS data to identify changes in forest land cover. The method successfully identified pixels affected by logging and fire disturbance in temperate and boreal forest sites.

Preparation of a Block of Aerial Photography Using the Leica Photogrammetry Suite®

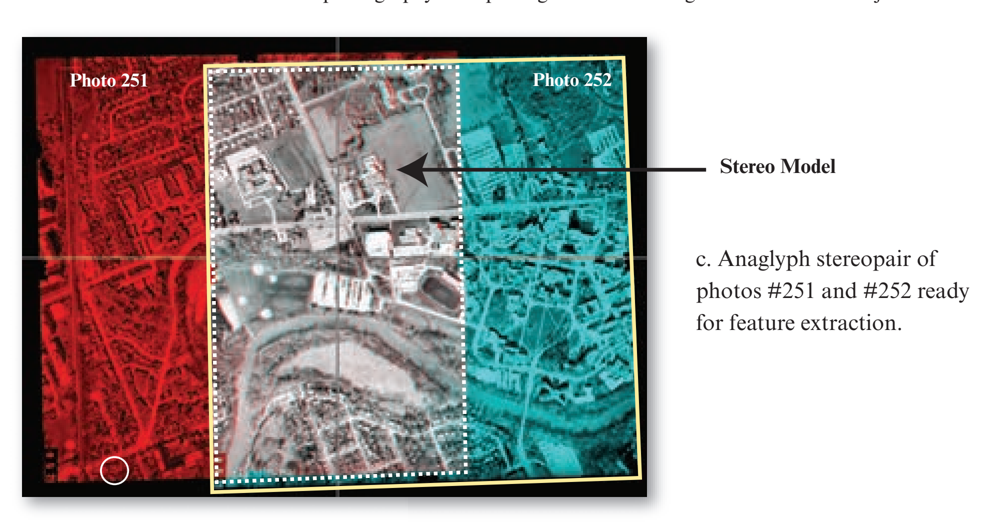
Photo 251 Photo 252 Photo 253

Display Mode
● Map Space
○ Image Space
Image Extents
Image IDs
Control Points
Tie Points
Check Points
Point IDs
Residuals

a. Distribution of ground control and tie points in a block of three aerial photographs (#251, #252, and #253).

Row #	Image ID	Description	Image Name	Active	Pyramid	Interior	Exterior	DTM	Orthophoto
1	1		251.img	●		complete			
2	2		252.img	●					
3	3		253.img	●					

b. Status of the block of photography after photogrammetric triangulation and block adjustment.



c. Anaglyph stereopair of photos #251 and #252 ready for feature extraction.

FIGURE 12–21 a) Control associated with a block of three vertical aerial photographs. b) Status of the preparation of the block of photography. c) Anaglyph stereo pair of photos #251 and #252 with the stereo model annotated (analyzed using Leica Photogrammetry Suitefi software).



Thematic "From–To" Change Detection Algorithms

While binary "change versus no-change" information is sufficient for many applications, it is much better to collect information on what the land use or land cover was and what it turned into. This is commonly referred to as obtaining thematic "from–to" change detection information. There are a variety of procedures and algorithms that can be used to extract "from–to" information from multiple-date remote sensor data. We will focus our attention on photogrammetric, LiDAR-grammetric, Post-Classification Comparison change detection using both per-pixel and object-based (OBIA) classifications, Neighborhood Correlation Image (NCI) Change Detection, and Spectral Change Vector Analysis change detection.

Photogrammetric Change Detection

Commercial photogrammetric engineering firms routinely collect high spatial resolution stereoscopic aerial photography for public and commercial applications. A

Feature Extraction of Buildings and Road using ERDAS Stereo Analyst



b. Feature extraction using the floating mark (circled in yellow) while viewing the anaglyph stereo model of photos #251 and #252.

c. Buildings and road shapefiles extracted using soft-copy photogrammetry draped over photographs #251 and #252.

FIGURE 12–22 a) A properly prepared stereo model can be analyzed using soft-copy photogrammetric techniques to extract very accurate building and road network information. Here only a small part of the stereo model is being analyzed. b) The operator uses the floating mark to map building footprints and roads. c) Overlay of the shapefile on top of the two aerial photographs (analyzed using ERDAS Stereo Analyst® software).

calibrated digital frame camera is typically used to collect the panchromatic, color and/or color-infrared digital aerial photography. The photography is collected in systematic flightlines that usually have 60% to 80% endlap and 20% sidelap (Jensen, 2007). Satellites such as WorldView-2 also collect stereoscopic imagery.

Horizontal and/or vertical ground control points (GCPs) are obtained on the ground using *in situ* surveying in support of the data collection. These GCPs, additional image-derived tiepoints, and onboard GPS/inertial navigation information are used by the analyst to perform interior and exterior orientation to scale and level the stereoscopic models. Often, three-dimensional LiDAR data are used for control (Mitishita et al., 2008). The digital aerial photography, camera calibration information, and ground control are subjected to a rigorous aero-triangulation procedure that prepares the block of aerial photography for information extraction. For example, the block of three vertical aerial photographs (#251, #252, and #253) shown in Figure 12-21a was prepared for stereoscopic analysis using interior and exterior orientation and aero-triangulation (Figure 12-21b). Each of the resultant stereoscopic models) can be analyzed in three dimensions using anaglyphic glasses. The stereo pair associated with photographs #251 and #252 is shown in Figure 12-21c. Note the 60% overlap area that constitutes the stereoscopic model. Properly prepared stereoscopic aerial photography can be analyzed using photogrammetric mapping techniques to extract building footprints, road networks, vegetation, water distribution, etc. on individual dates of imagery (Elaksher, et al., 2003). For example, consider the small part of the stereo pair (photos #251 and #252) shown in Figure 12-22a that consists of single-family residential housing. The two photographs in the stereo model are projected onto the screen, one in blue and one in red. The analyst views the stereo model using anaglyph glasses that have corresponding blue and green lenses. The mind fuses the two views of the terrain and creates a three-dimensional view of the terrain on the computer screen (Figure 12-22b). The analyst then uses a three-dimensional "floating mark" to identify features of interest (within the yellow circle in Figure 12-22b).

If a building footprint is desired, the analyst raises the floating mark above the terrain to the elevation of a vertex on the edge of the building roof. The floating mark is then moved around the perimeter of the roof to important vertices where data are acquired. Upon completion, the polygon is closed and represents the building footprint as if it were measured on the ground by a surveyor. Road network information is obtained by simply moving the floating mark so that it rests on the ground (i.e., ground elevation) and then clicking on points along the edge of the road (or the centerline of the road) while keeping the floating mark on the ground. Any



FIGURE 12–23 a) Date 1 aerial photography with the building footprint extracted and outlined in yellow. The footprint of the same building derived from the Date 2 photography is displayed in red. b) Date 2 aerial photograph with the building footprint extracted displayed in red (aerial photography courtesy of Pictometry International, Inc.).

feature in the scene such as hydrographic features, driveways, and sidewalks can be mapped. All these geospatial features of interest are mapping in their exact planimetric (*x*, *y*) position (Figure 12-22c).

Photogrammetric compilation is performed on the aerial photography obtained on Date 1. Thematic information such as buildings and roads that have not changed do not have to be recompiled on Date 2 aerial photography. Rather, only the new structures or roads are extracted in the Date 2 photography and added to the database. For example, the Pictometry International (2014) ChangeFindrfi program automatically identifies changes in building footprints as demonstrated in Figure 12-23. Such information can be used to determine if unauthorized building additions have been made. If desired, additional information can be extracted about other features such as the change from dirt to paved driveway and landscape modification. This type of three-dimensional data are ideal for subsequent use in urban change detection projects, e.g., to identify the change in city 3D building models (Qin, 2014).

LiDARgrammetric Change Detection

Commercial photogrammetric engineering firms also routinely collect high spatial resolution Light Detection and Ranging (LiDAR) data for a variety of commercial and public applications (Figure 12-24). The LiDAR data are acquired using special instruments that transmit and receive hundreds of thousands of near-infrared (1,024 nm) pulses of laser energy every second (Jensen, 2007). Normally, only a few well-placed horizontal and/or vertical ground control points (GCPs) obtained using *in situ* surveying are required when collecting LiDAR data for a given region. These GCPs and the onboard GPS/inertial navigation information are used to accurately locate the *x*, *y*, and *z*-position of every LiDAR first return, intermediate return, last return, and intensity return. These data are used to prepare digital surface models (DSMs) of the terrain that contain *x*, *y*, and *z* information about every tree, bush, building and road. These data can also be processed to prepare digital terrain models (DTMs) that contain only the bare Earth digital elevation information (Jensen, 2007).

Properly prepared DSMs and DTMs can also be used to identify the location of individual buildings and road network using the individual Date 1 and Date 2 LiDAR datasets. Building footprints and road networks extracted on one date will be in the exact planimetric location on the next date if the features have not changed. Buildings and roads or other features carefully extracted from Date 1 LiDAR data can be compared with buildings and roads and other features extracted from the Date 2 LiDAR data to identify changes (e.g., Sohn and Dowman, 2007).



a. Building footprints for 33 cities and towns in the metropolitan Boston area.

b. Enlargement of a small part of the metropolitan Boston study area.

FIGURE 12–24 a) LiDAR-derived building footprints for 33 cities and towns in metropolitan Boston, MA. The buildings are so dense in most areas that they appear dark gray at this small scale. b) An enlargement of a small part of the study area. Such LiDAR-derived information obtained on multiple dates is used routinely to monitor change (source of data: www.mass.gov/mgis/massgis.htm).

Many cities, counties, and state departments of natural resources now routinely collect LiDAR data on a repetitive basis and extract information from the data. For example, the Massachusetts Office of Geographic Information (MassGIS) extracted all of the building footprints for approximately 33 cities and towns in the metropolitan Boston area in 2002 (Figure 12-24a). An enlargement of a small part of the study area is shown in Figure 12-24b. It is now quite common for LiDAR-derived geospatial information to be used for change detection purposes.

Post-Classification Comparison Change Detection

Post-classification comparison change detection is one of the most widely used change detection methods (e.g., Ahlqvist, 2008; Warner et al., 2009; Griffiths et al., 2010; Tsai et al., 2011; Rokni et al., 2015). The method requires accurate geometric rectification of the two dates of imagery to a common map projection to within RMSE $\pm$0.5 pixel. The Date 1 image and the Date 2 image are classified using a) per-pixel analysis, or b) object-based image analysis (OBIA) (Gladstone et al., 2012). The classification schemes used for the Date 1 and Date 2 classification should be identical. A thematic change map is produced using a change detection matrix based on the use of a simple GIS overlay function (van Oort, 2007; Taubenbock et al., 2012).

For example, the change detection matrix logic shown in Figure 12-25 can be used to highlight important change classes in the change thematic map, e.g., all of the pixels in the change image with a value of 9 were vegetation in the Date 1 classification and are now buildings in the Date 2 classification. Unfortunately, every classification error in the Date 1 and Date 2 classification maps will also be present in the final change detection thematic map (Rutchey and Velcheck, 1994; Jensen et al., 1995). The minimum amount of change error in the change map is equivalent to the less accurate of the two individual classifications (van Oort, 2007; Warner et al., 2009). Therefore, it is imperative that the individual date classification maps used in the post-classification change detection be as accurate as possible (Purkis and Klemas, 2011).

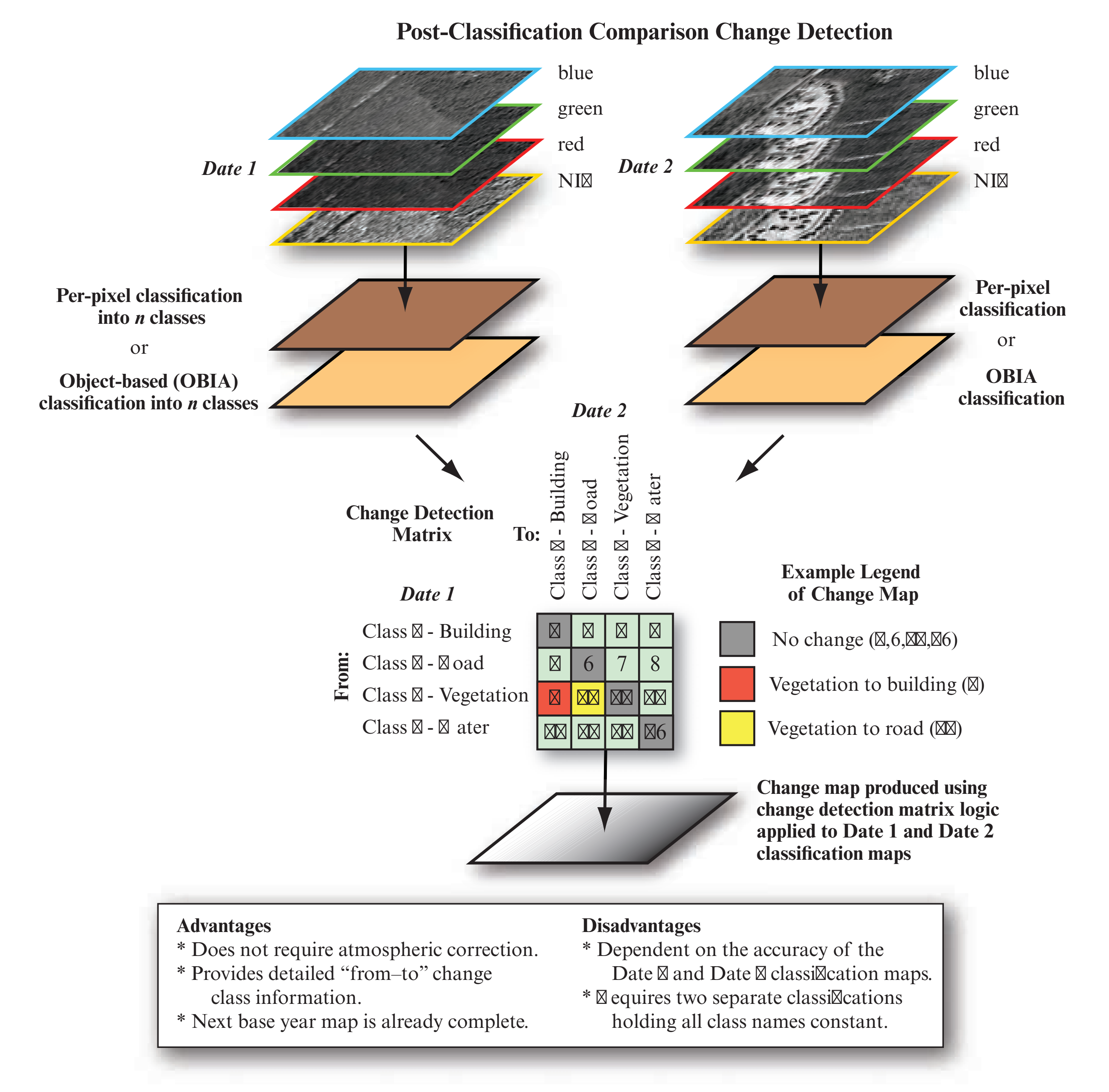


FIGURE 12–25 The logic of performing post-classification comparison change detection using multiple dates of imagery classified using either a per-pixel or object-based image analysis (OBIA) classification and a change detection matrix tailored to meet the change detection information requirements (imagery courtesy of Richland County GIS Department).

Per-Pixel Post-Classification Comparison

An example of post-classification comparison change detection applied to per pixel classification is shown in Figures 12-26 and 12-27 using Landsat Thematic imagery of an area centered on Fort Moultrie, SC, near Charleston. Nine land-cover classes were inventoried on each date (Figure 12-26). The 1982 and 1988 classification maps were then compared on a pixel-by-pixel basis using an n ⊠ n GIS matrix algorithm whose logic is shown in Figure 12-27b. This resulted in the creation of a change image map consisting of brightness values from 1 to 81. The analyst then selected specific "from–to" classes for emphasis. Only a select number of the 72 possible off-diagonal "from–to" land-cover change classes summarized in the change matrix (Figure 12-27b) were selected to produce the change detection map shown in Figure 12-27a. For example, all pixels that changed from any land cover in 1982 to Developed Land in 1988 were color coded red (RGB = 255, 0, 0) by selecting the appropriate "from–to" cells in the change detection matrix (10, 19, 28, 37, 46, 55, 64, and 73). Note that the change classes are draped over a Landsat TM band 4 image of the study area to facilitate orientation. Similarly, all pixels in 1982 that changed to Estuarine Unconsolidated Shore by December 19, 1988 (cells 9, 18, 27, 36, 45, 54, 63, and 72),

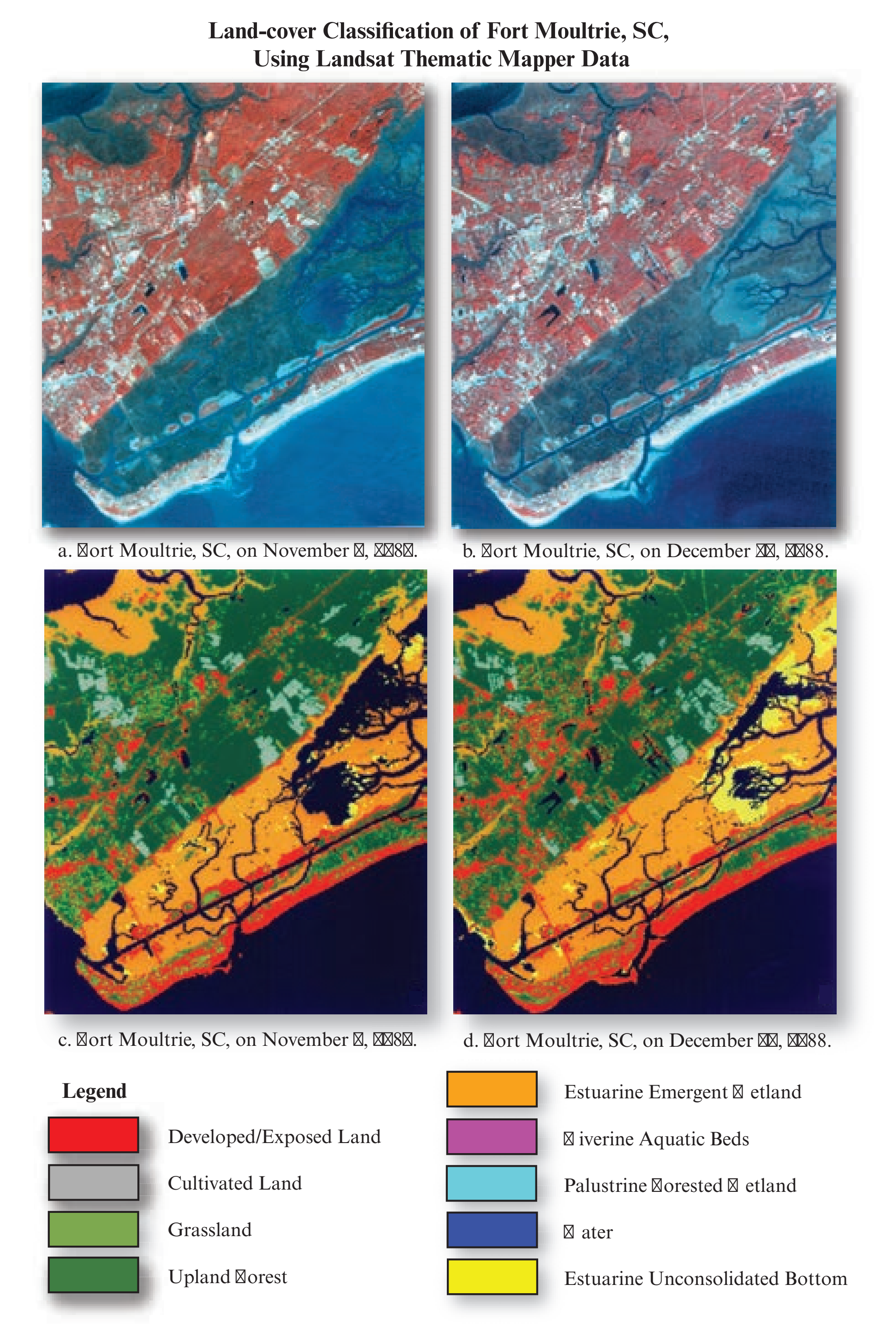


FIGURE 12–26 ab) Rectified Landsat Thematic Mapper data of an area centered on Fort Moultrie, SC, obtained on November 9, 1982, and December 19, 1988 (RGB = bands 4, 3, 2). cd) Classification maps of the Fort Moultrie, SC, study area produced from the November 9, 1982, and December 19, 1988, Landsat TM data. Some barren land is included in the Developed/Exposed Land category (original imagery courtesy of NASA).

were depicted in yellow (RGB = 255, 255, 0). If desired, the analyst could highlight very specific changes such as all pixels that changed from Developed Land to Estuarine Emergent Wetland (cell 5 in the matrix) by assigning a unique color look-up table value (not shown). A color-coded version of the change detection matrix can be used as an effective "from–to" change detection map legend.

OBIA Post-Classification Comparison

Object-based image analysis (OBIA) can extract land-cover information on individual dates of high spatial

Selected Change in Land Cover for Fort Moultrie, SC

Change in Land Cover from November 9, 1982, to December 19, 1988

New Developed/Exposed Land
New Estuarine Emergent Wetland
New Riverine Aquatic Beds
New Palustrine Forested Wetland
New Water
New Estuarine Unconsolidated Bottom
No change in land cover between dates, and not selected for display (RGB = 0, 0, 0).

a. Fort Moultrie, SC.

b.

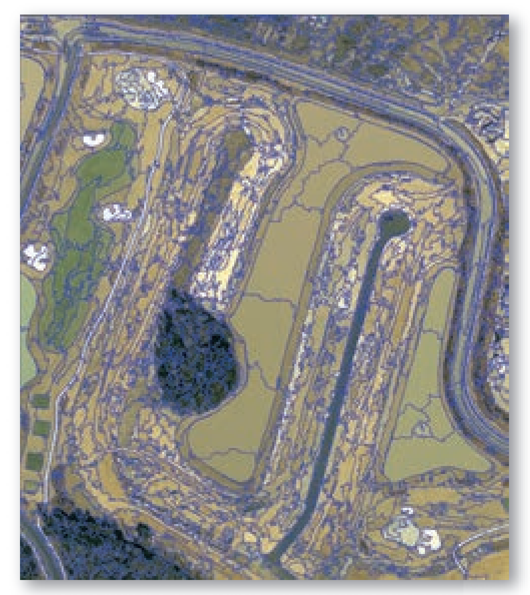
***From – To* Change Detection Matrix**

From 1982 \ To 1988	Developed/Exposed Land	Cultivated Land	Grassland	Upland Forest	Estuarine Emergent Wetland	Riverine Aquatic Beds	Palustrine Forested Wetland	Water	Estuarine Unconsolidated Bottom
Developed/Exposed Land	1	2	3	4	5	6	7	8	9
Cultivated Land	10	11	12	13	14	15	16	17	18
Grassland	19	20	21	22	23	24	25	26	27
Upland Forest	28	29	30	31	32	33	34	35	36
Estuarine Emergent Wetland	37	38	39	40	41	42	43	44	45
Riverine Aquatic Beds	46	47	48	49	50	51	52	53	54
Palustrine Forested Wetland	55	56	57	58	59	60	61	62	63
Water	64	65	66	67	68	69	70	71	72
Estuarine Unconsolidated Bottom	73	74	75	76	77	78	79	80	81
Color look-up table values in change detection map Red	255	255	255	255	255	255	0	0	255
Green	0	255	255	255	163	0	255	0	255
Blue	0	255	255	255	0	255	255	255	0

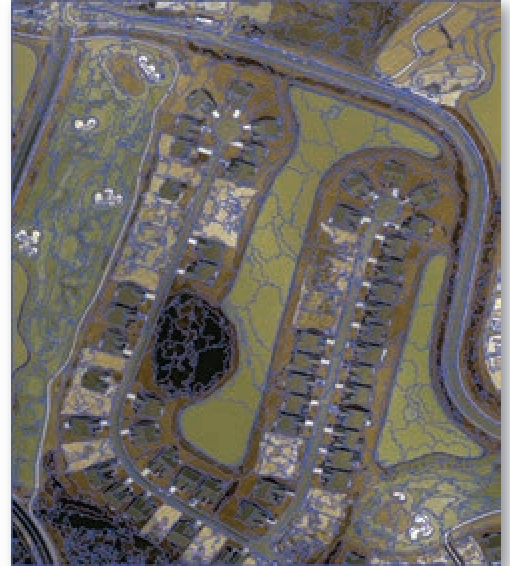
FIGURE 12–27 a) Change detection map of the Fort Moultrie, SC, study areas derived from analysis of November 11, 1982, and December 19, 1988, Landsat Thematic Mapper data. The natures of the change classes selected for display are summarized in b) the "From–To" Change Detection Matrix (based on Jensen et al., 1993a).

resolution imagery (Figure 12-28) (e.g. Zhan et al., 2005; Hofmann et al., 2008; Stow et al., 2011). Hurskainen and Pellikka (2004) used OBIA to study the growth and change of informal settlements in Southeast Kenya using scanned aerial photography from three different dates. The post-classification comparison change detection method yielded high accuracy detecting new buildings. Moeller and Blaschke (2006) used multiple-date QuickBird imagery to detect new buildings in Phoenix, AZ. Matikainen et al. (2010) used LiDAR data and identified building changes using an automated OBIA approach. They concluded

Post-classification Comparison Change Detection using Object-Based Image Analysis (OBIA)



a. 2007 Bluffton, SC, 1 × 1 ft. imagery segmented to 40 scale.



b. 2011 Bluffton, SC, 1 × 1 ft. imagery segmented to 40 scale.



c. 2007 imagery segmented to 120 scale.

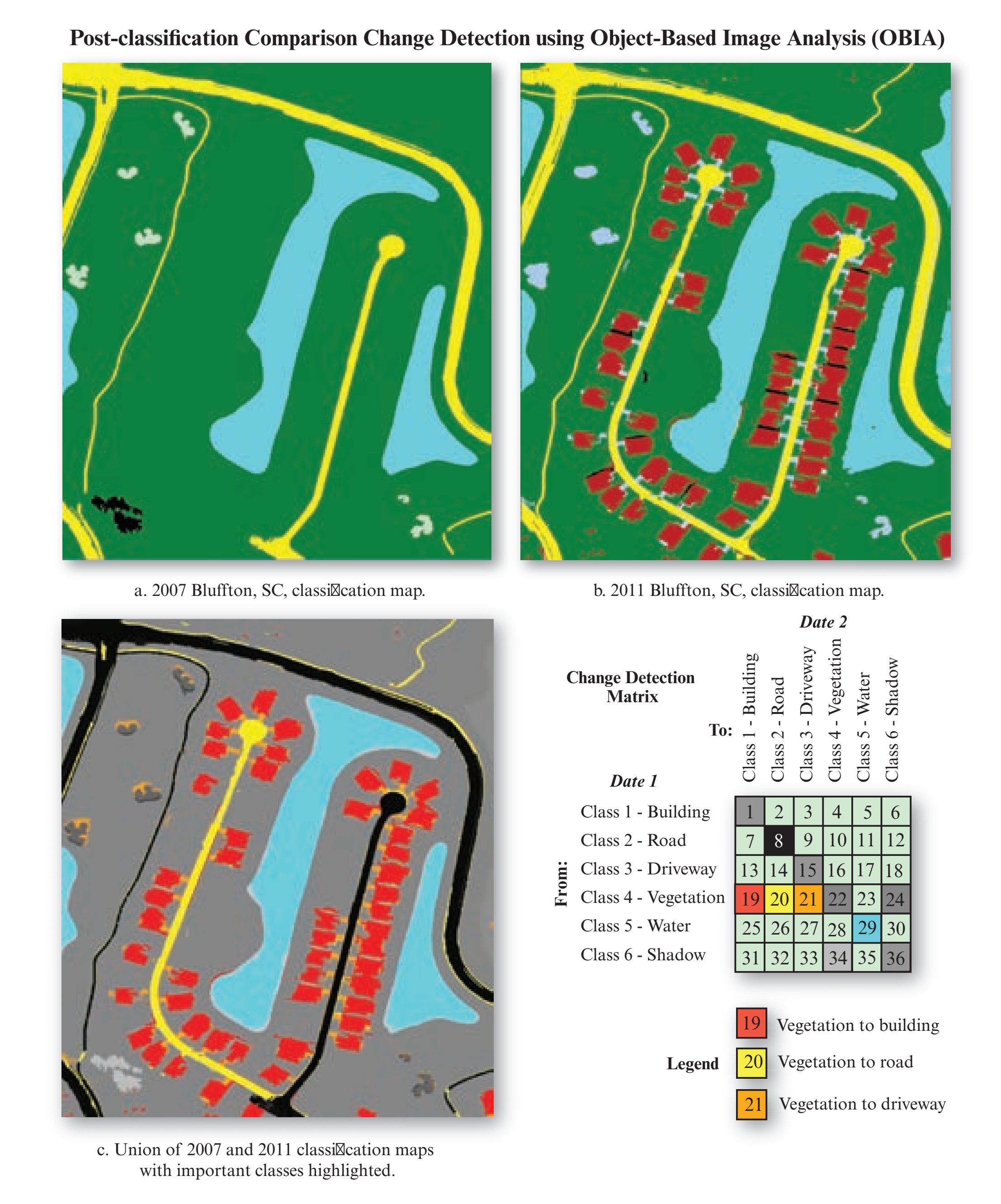


d. 2011 imagery segmented to 120 scale.

FIGURE 12–28 Post-classification comparison can be performed using per-pixel or object-based image analysis classification. This is an example of OBIA segmentation. Selected object-based segmentation of the Bluffton, SC, Pictometry International, 2007 and 2011 natural color aerial photography at (a,b) 40 scale, and (c,d) 120 scale (imagery courtesy of Beaufort County GIS Department).

that the main problem of building change detection was false detection due to new or demolished buildings, and confusion with adjacent (background) features. Tsai et al. (2011) delineated new buildings by combining OBIA approaches using multi-date high spatial resolution satellite data. Buildings constructed between 2002 and 2010 in Accra, Ghana, were delineated and quantified. Taubenbock et al. (2012) used decision-tree classification and post-classification comparison to detect change in very large cities (>10 million people). Chen et al. (2012) provided an overview of object-based change detection logic and considerations. Gartner et al. (2014) used OBIA to extract tree crown area and diameter in multiple dates of satellite imagery and then compared the diameter of the crowns to detect change.

An example of post-classification comparison change detection using OBIA applied to 2007 and 2011 Pictometry, Inc., 1 × 1 ft. spatial resolution digital aerial photography of an area in Bluffton, SC, is shown in Figures 12-28 and 12-29. The 2007 and 2011 aerial



a. 2007 Bluffton, SC, classification map.

b. 2011 Bluffton, SC, classification map.

c. Union of 2007 and 2011 classification maps with important classes highlighted.

FIGURE 12–29 Post-classification change detection of Bluffton, SC, based on 2007 and 2011 Pictometry, Inc., natural color aerial photography and object-based image analysis. a) Classification of 2007 photography (rasterized). b) Classification of 2011 photography (rasterized). c) Union of 2007 and 2011 classification maps using GIS map overlay function with selected "from–to" classes highlighted.

photography was segmented to the following "scales" using *e*Cognition OBIA: 40, 50, 60, 70, 80, 90, 100, 110, and 120. Only the 40- and 120-scale segmentations are shown in Figures 12-28ab and 12-28cd, respectively. The 2007 aerial photography was classified using the coarse 120 scale segmentation level because this level was sufficient to capture the spatial and spectral detail of the land cover present on this date (Figure 12-29a). Conversely, the 2011 photography was classified using the more detailed 40-scale segmentation because it was necessary to identify the vegetation and shadows between the individual residential houses (Figure 12-29b).

The post-classification change detection map was created by analyzing the 2007 and 2011 OBIA classification maps (converted to raster format) using a GIS overlay "union" function, resulting in 36 possible "from–to" change classes shown in Figure 12-29. Several important "from–to" change classes are highlighted in Figure 12-29c: from vegetation to building (red), from vegetation to road (yellow), and from vegetation

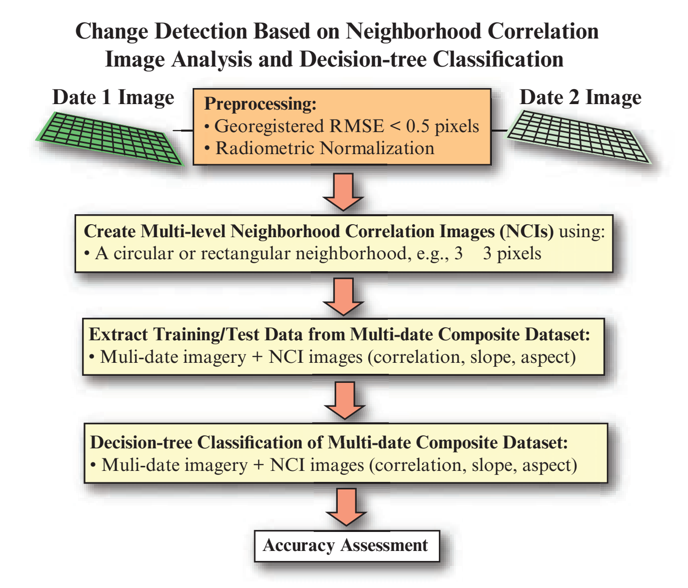


FIGURE 12–30 Steps required to perform change detection using Neighborhood Correlation Image (NCI) analysis and decision-tree classification (based on Figure 2 in Im and Jensen (2005) cited in Figure 12-31).

to driveway (orange). Roads and water present in both dates of photography are shown in black and cyan, respectively, for illustrative purposes. Note that the geometric registration between dates was very good as the narrow golf cart paths and the major roads exhibited only a slight amount of error along the edges.

Post-classification comparison change detection is widely used and easy to understand. When conducted by skilled image analysts, it represents a viable technique for the creation of change detection products. Advantages include the detailed "from–to" information that can be extracted and the fact that the classification map for the next base year is already complete (Arzandeh and Wang, 2003; Jensen et al., 2012). However, the accuracy of the change detection depends on the accuracy of the two separate classification maps.

Neighborhood Correlation Image (NCI) Change Detection

Neighborhood Correlation Image (NCI) Change Detection is based on the creation of a new three-channel NCI that contains three unique types of information (Im and Jensen, 2005; Im et al., 2007, 2008):

- correlation,
- slope, and
- intercept

that can be related to the change in multiple-date imagery (Figure 12-30). The correlation information derived from a specific neighborhood of pixels (e.g., 3 × 3) contains valuable change information associated with a central pixel and its contextual neighbors. Slope and intercept images provide change-related information that can be used as ancillary data to facilitate change detection with correlation. The degree of correlation, slope, and intercept can then be used to produce detailed "from–to" change information when combined with expert system (e.g., decision tree) or object-oriented classification techniques.

NCI analysis is based on the change magnitude and direction of brightness values by band in a specific neighborhood between two multispectral remote sensing datasets. If the spectral changes of the pixels within a specified neighborhood between two image dates are significant, the correlation coefficient between the two sets of brightness values in the neighborhood will fall off to a lower value. The slope and intercept values may increase or decrease depending on the magnitude and direction of the spectral changes. Ideally, if there is no change, pairs of brightness values between the two dates of imagery should be located along y = x, and be located far from y = x if change has occurred between the two image dates. Consequently, correlation coefficients are lower when change occurs and higher when only a small amount of change is present. Intermediate correlation values (e.g., r = 0.7) are generally encountered in areas that have changed only slightly (e.g., widened roads, expanded buildings) or changed vegetation status (e.g., the same vegetation captured in a slightly different phenological cycle).

Correlation analysis is used to detect change in relatively small local neighborhoods (e.g., 3 × 3 pixels) in two registered remote sensor images obtained on different dates. The correlation between the brightness values of two dates of imagery in the local neighborhoods is calculated using the equations (Im and Jensen, 2005; Im et al., 2007):

$$r = \frac{cov_{12}}{s_1 s_2} \tag{12.6}$$

where

$$cov_{12} = \frac{\sum_{i=1}^{n} (BV_{i1} - \mu_1)(BV_{i2} - \mu_2)}{n - 1} \tag{12.7}$$

and r is Pearson's product-moment correlation coefficient, cov_{12} is the covariance between brightness values found in all bands of the two date datasets in the neighborhood, and s_1 and s_2 are the standard deviations of the brightness values found in all bands of each dataset in the neighborhood, respectively. BV_{i1} is the ith brightness value of the pixels found in all bands of image 1 in the neighborhood, BV_{i2} is the ith brightness value of the pixels found in all bands of image 2 in the neighborhood, n is the total number of the pixels

Images used in Neighborhood Correlation Image Analysis Change Detection



a. Near-infrared band obtained on Sept. ⊠⊠, ⊠⊠⊠⊠.



b. Near-infrared band obtained on Oct. ⊠⊠, ⊠⊠⊠⊠.

FIGURE 12–31 Near-anniversary multiple date (1999 and 2000) near-infrared imagery of an area in Edisto Beach, SC, at 1 ⊠ 1 ft. spatial resolution. New residential housing that is completely built is highlighted in red. New cleared land is highlighted in yellow (based on Figure 1 in Im, J., and J. R. Jensen, 2005, "Change Detection Using Correlation Analysis and Decision Tree Classification," *Remote Sensing of Environment*, 99:326–340).

found in all bands of each dataset in the neighborhood, and μ_1 and μ_2 are the means of brightness values found in all bands of the two date (1 and 2) images in the neighborhood, respectively.

Slope (a) and y-intercept (b) are calculated using the following equations from the least squares estimates:

$$a = \frac{cov_{12}}{s_1^2} \quad (12.8)$$

$$b = \frac{\sum_{i=1}^{n} BV_{i2} - a\sum_{i=1}^{n} BV_{i1}}{n} \quad (12.9)$$

Figure 12-30 summarizes the steps required to implement the change detection based on neighborhood correlation image analysis and decision tree classification, including:

- Geometrically and radiometrically preprocess the multi-temporal remote sensor data (previously discussed).
- Create the multi-level Neighborhood Correlation Images (NCIs), i.e., correlation, slope, and intercept images (bands) based on the neighborhood configuration specified.
- Collect training/test data from the multi-date, multi-channel dataset (e.g., eight spectral bands plus multi-level NCI features: correlation, slope, and intercept).
- Use the training/test data to generate the production rules to extract "from–to" change information based on decision-tree logic classification and create a change detection map.
- Perform an accuracy assessment to determine the effectiveness of the change detection model.

Digital frame camera imagery of a residential area in Edisto Beach near Charleston, SC, is used to demonstrate the utility of Neighborhood Correlation Images for change detection (Figure 12-31ab). The data were obtained on September 23, 1999, and October 10, 2000, in four spectral bands (blue, green, red, and near-infrared). Both dates of imagery were geo-referenced to a Universal Transverse Mercator projection with an RMSE < 0.5 pixels. The data were radiometrically normalized using water and bare soil radiometric ground control points identified as pseudo-invariant pixels using techniques described in Chapter 6. Regression equations allowed the individual September 23, 1999, multispectral bands to be radiometrically normalized to the October 10, 2000, multispectral bands. Correlation, slope, and intercept images derived from the multi-date images are displayed in Figure 12-32.

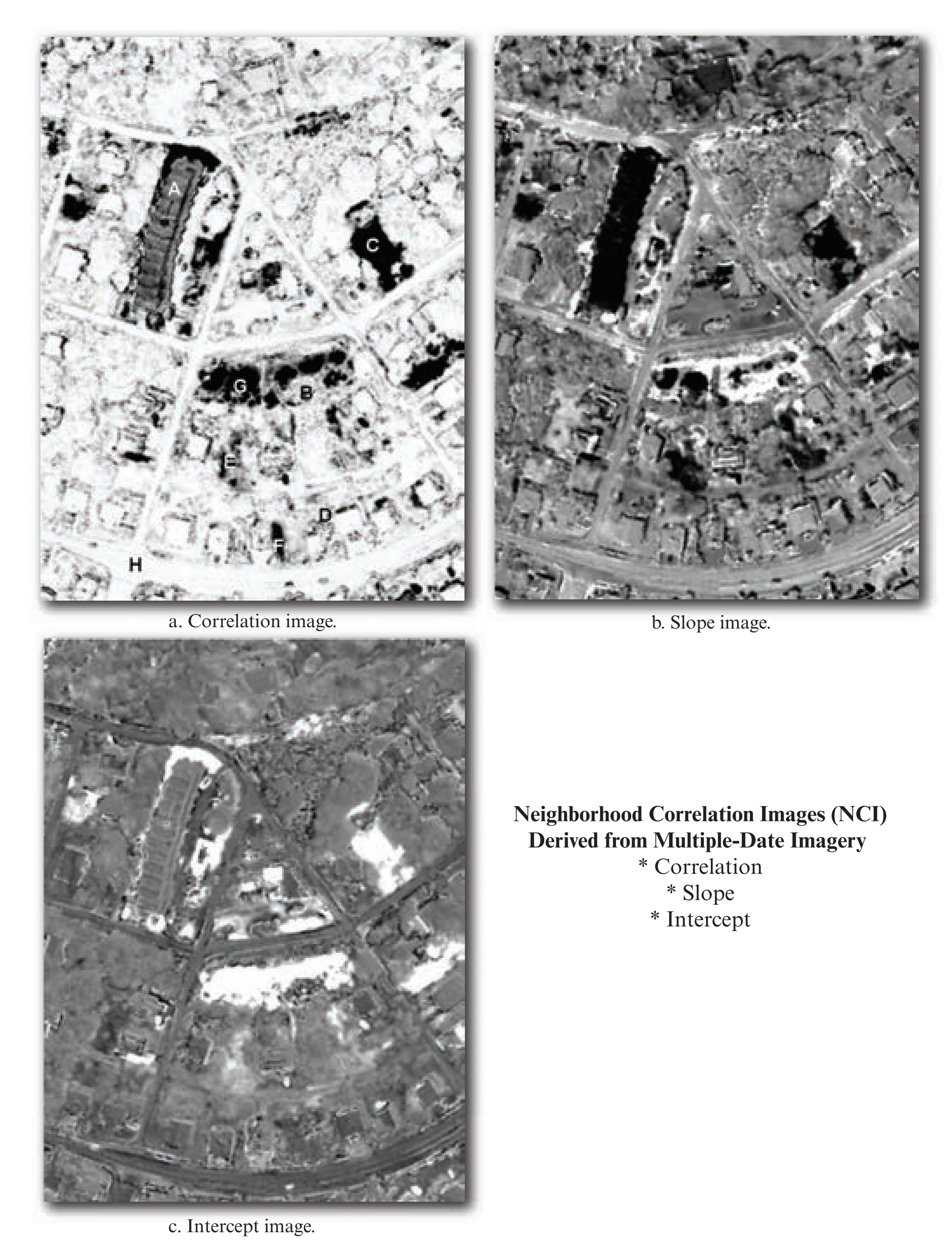


FIGURE 12–32 Three Neighborhood Correlation Images (NCI) derived from multiple-date imagery including a) correlation, b) slope, and c) intercept (based on Figure 4 in Im, J., and J. R. Jensen, 2005, "Change Detection Using Correlation Analysis and Decision Tree Classification," *Remote Sensing of Environment*, 99:326–340).

The correlation coefficient (r) indicates the spectral change between the two images on a pixel by pixel basis. However, high correlation does not always guarantee that change has not occurred. For example, Figure 12-33 displays the results of examining just three locations within the 1999 and 2000 high spatial resolution images. Sample A is from an unchanged area (Developed on both dates). Samples B and C are from changed areas (grass to barren, and barren to developed, respectively). All three locations exhibit high correlation (i.e., $r > 0.8$). Thus, locations B and C are not separable from the unchanged area based solely on correlation coefficient information with a threshold of $r = 0.8$. However, when we consider the slope and y-intercept of the regression lines shown in Figure 12-33, we find that B and C are from changed areas. Ideally, a pixel from an unchanged area will have a slope of ~1 and a y-intercept near 0 similar to sample A shown in

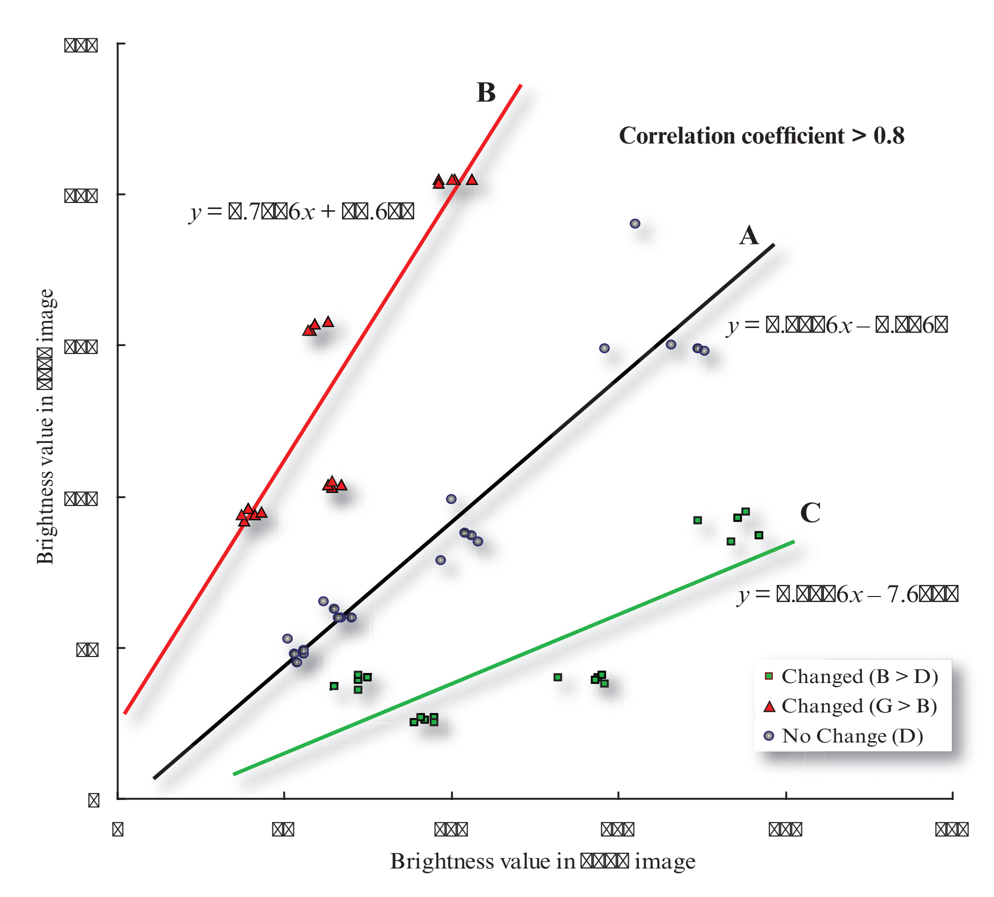


FIGURE 12–33 Correlation results associated with the Edisto, SC, multiple-date (1999 and 2000) imagery (based on Figure 3 in Im, J., and J. R. Jensen, 2005, "Change Detection Using Correlation Analysis and Decision Tree Classification," *Remote Sensing of Environment*, 99:326–340).

Change Detection Based on Correlation Image Analysis and Decision-Tree Classification

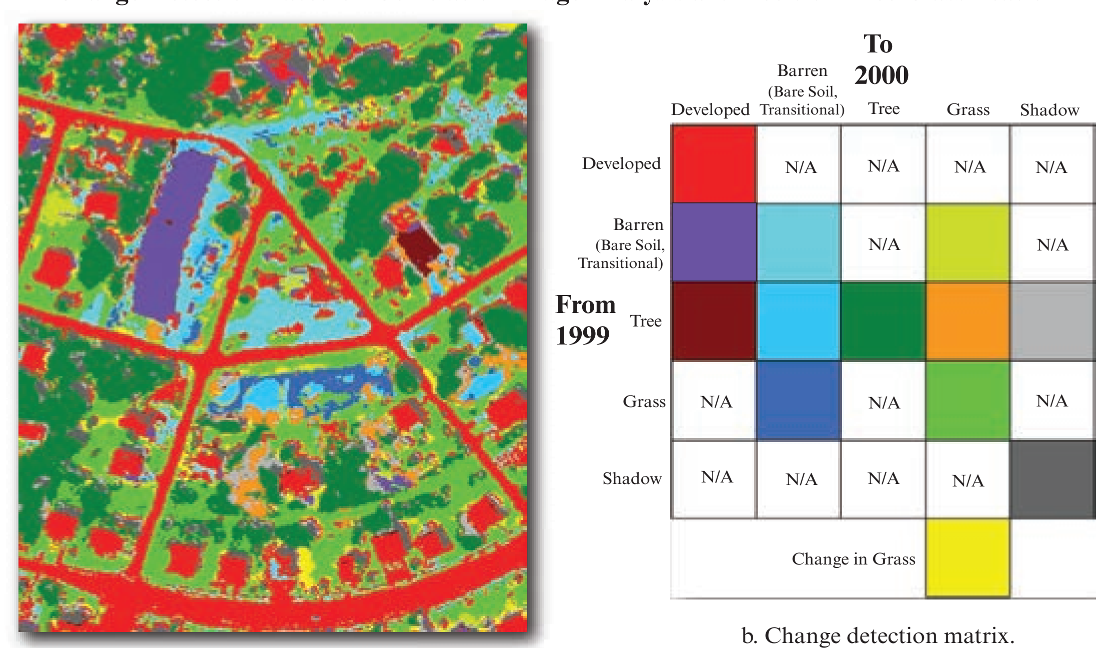


FIGURE 12–34 "From-to" change information based on Neighborhood Correlation Image (NCI) analysis and decision-tree classification (based on Figure 8 in Im, J., and J. R. Jensen, 2005, "Change Detection Using Correlation Analysis and Decision Tree Classification," *Remote Sensing of Environment*, 99:326–340).

FIGURE 12–35 a-d) Schematic diagram of the Spectral Change Vector Analysis change detection method. e) Possible change sector codes for a pixel measured in three bands on two dates.

Figure 12-33. Conversely, change areas like B and C have relatively higher or lower slopes and *y*-intercepts. In this example, location B has a slope of ~1.8 and its *y*-intercept is ~25. Location C has a slope of < 0.5 and a negative *y*-intercept. Consequently, locations A, B, and C can be distinguished as being change or no-change by incorporating slope and intercept data in addition to correlation information.

Decision-tree logic can be used to classify the multi-date composite image datasets consisting of two dates of image data plus the three NCI features (correlation, slope, and intercept). Rules generated from a C5.0 decision tree were used in an expert system to identify "from–to" change information on a pixel by pixel basis (Figure 12-34).

Spectral Change Vector Analysis

When land undergoes a change or disturbance between two dates, its spectral appearance normally changes. For example, consider the red and near-infrared spectral reflectance characteristics of a single pixel displayed in two-dimensional feature space (Figure 12-35a). It appears that the land cover associated with this particular pixel has changed from Date 1 to Date 2 because the pixel resides at a substantially different location in the feature space on Date 2. The vector describing the direction and magnitude of change from Date 1 to Date 2 is a *spectral change vector* (Malila, 1980; Warner et al., 2009; Carvalho et al., 2011; Kontoes, 2013; Yuan et al., 2015).

The total change magnitude per pixel (CM_{pixel}) is computed by determining the Euclidean distance between endpoints through *n*-dimensional change space (Michalek et al., 1993; Kontoes, 2013):

$$CM_{pixel} = \sum_{k=1}^{n} [BV_{ijk(\text{date2})} - BV_{ijk(\text{date1})}]^2 \quad (12.10)$$

where $BV_{i,j,k(\text{date2})}$ and $BV_{i,j,k(\text{date1})}$ are the Date 1 and Date 2 pixel values in band k.

A scale factor (e.g., 5) can be applied to each band to magnify small changes in the data if desired. The change direction for each pixel is specified by whether the change is positive or negative in each band. Thus, 2^n possible types of changes can be determined per pixel (Virag and Colwell, 1987). For example, if three bands are used there are 2^3 or 8 types of changes or sector codes possible as shown in Table 12-4. To demonstrate, consider a single registered pixel measured in just three bands (1, 2, and 3) on two dates. If the change in band 1 was positive (e.g., $BV_{i,j,1(\text{date 2})} = 45$; $BV_{i,j,1(\text{date 1})} = 38$; $BV_{\text{change}} = 45 - 38 = 7$), and the change in band 2 was positive (e.g., $BV_{i,j,2(\text{date2})} = 20$;

TABLE 12–4 Sector code definitions for change vector analysis using three bands (+ indicates pixel value increase from Date 1 to Date 2; – indicates pixel value decrease from Date 1 to Date 2) (Michalek et al., 1993).

	Change Detection		
Sector Code	Band 1	Band 2	Band 3
1	–	–	–
2	–	–	+
3	–	+	–
4	–	+	+
5	+	–	–
6	+	–	+
7	+	+	–
8	+	+	+

$BV_{i,j,2(\text{date1})} = 10$; $BV_{\text{change}} = 20 - 10 = 10$), and the change in band 3 was negative (e.g., $BV_{i,j,3(\text{date2})} = 25$; $BV_{i,j,3(\text{date1})} = 30$; $BV_{\text{change}} = 25 - 30 = -5$), then the change magnitude of the pixel would be $\text{CM}_{\text{pixel}} = 7^2 + 10^2 - 5^2 = 174$, and the change sector code for this pixel would be "+, +, –" and have a value of 7, as shown in Table 12-4 and Figure 12-35e. For rare instances when pixel values do not change at all between the two dates, a default direction of + may be used to ensure that all pixels are assigned a direction (Michalek et al., 1993).

Change vector analysis outputs two geometrically registered files; one contains the sector code and the other contains the scaled vector magnitudes. The change information may be superimposed onto an image of the study area with the change pixels color-coded according to their sector code. This multispectral change magnitude image incorporates both the change magnitude and direction information (Figure 12-35a). The decision that a change has occurred is made if a threshold is exceeded (Virag and Colwell, 1987). The threshold may be selected by examining deep-water areas (if present), which should be unchanged, and recording their scaled magnitudes from the change vector file. Figure 12-35b illustrates a case in which no change would be detected because the threshold is not exceeded. Conversely, change would be detected in Figures 12-35c and d because the threshold was exceeded. The other half of the information contained in the change vector, that is, its direction, is also shown in Figure 12-35c and d. Direction contains information about the type of change. For example, the direction of change due to clearing should be different from change due to regrowth of vegetation.

Change vector analysis has been applied successfully to forest change detection in northern Idaho (Malila, 1980), mangrove and reef ecosystems along the coast of the Dominican Republic (Michalek et al., 1993), and agriculture in Brazil (Carvalho et al., 2011). It is the change detection algorithm of choice for producing the MODIS Vegetative Cover Conversion (VCC) product being compiled on a global basis using 250 m surface reflectance data (Zhan et al., 2002). The method is based on measuring the change in reflectance in just two bands, red ($\Delta\rho_{red}$) and near-infrared ($\Delta\rho_{nir}$), between two dates and using this information to compute the change magnitude per pixel,

$$\text{CM}_{pixel} = \sqrt{(\Delta\rho_{\text{red}})^2 + (\Delta\rho_{\text{nir}})^2} \quad (12.11)$$

and change angle (θ):

$$\theta_{pixel} = \arctan\left(\frac{\Delta\rho_{\text{red}}}{\Delta\rho_{\text{nir}}}\right) \quad (12.12)$$

The change magnitude and angle information is then analyzed using decision-tree logic to identify important types of change in the multiple-date MODIS imagery (Zhan et al., 2002). Chen et al. (2003) developed an improved change vector analysis methodology that assists in the determination of the change magnitude and change direction thresholds when producing land-cover change maps. Carvalho et al. (2011) developed improved methods of computing the spectral direction. The chief advantage of this approach is that it generates a single image of change information insensitive to illumination variation. Kontoes (2013) developed an operational version of change vector analysis.

Change Detection Using an Ancillary Data Source as Date 1

Sometimes there exists a land-cover data source that may be used in place of a traditional remote sensing image in the change detection process. For example, the U.S. Fish and Wildlife Service conducted a National Wetland Inventory (NWI) of the U.S. at 1:24,000 scale. Much of these data have been digitized. Instead of using a remotely sensed image as Date 1 in a coastal change detection project, it is possible to substitute the digital NWI map of the region (Figure 12-36). In this case, the NWI map is recoded to be compatible with the classification scheme being used. Next, Date 2 of the analysis is classified and then compared on a pixel-by-pixel basis with the Date 1 information using post-classification comparison methods. Traditional "from–to" information can then be derived.

Change Detection Using an Ancillary Data Source as Date 1

Date 1
Ancillary data source, e.g., National Wetlands Inventory (NWI) map
Date 2
blue
green
red
NIR
Rectified remote sensor data
Classification map of Date 2
Classification map of Date 1
1) Perform post-classification comparison change detection.
or
2) Update Date 1 NWI map with Date 2 change information using GIS dominate function.

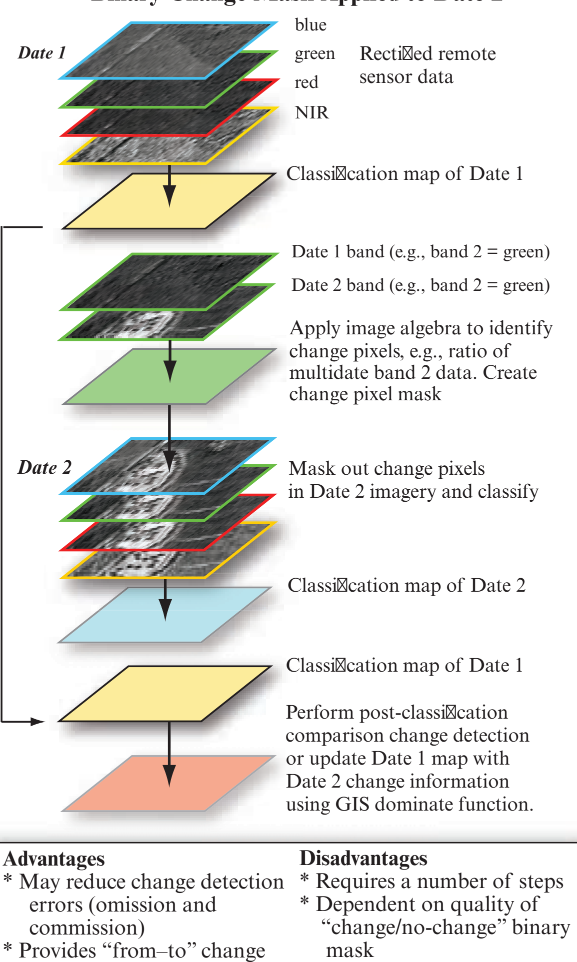
Advantages
* May reduce change detection errors (omission and commission)
* Provides "from–to" change class information
* Requires a single classification

Disadvantages
* Depends on quality of ancillary information

FIGURE 12–36 Diagram of change detection using an ancillary data source as Date 1 (imagery courtesy of Richland County GIS Department).

Advantages of the method include the use of a well-known, trusted data source (e.g., NWI) and the possible reduction of errors of omission and commission. Detailed "from–to" information may be obtained using this method. Also, only a single classification of the Date 2 image is required. It may also be possible to update the NWI map (Date 1) with more current wetland information (this would be done using a GIS dominate function and the new wetland information found in the Date 2 classification). The disadvantage is that the NWI data must be digitized, generalized to be compatible with a classification scheme, and then converted from vector to raster format to be compatible with the raster remote sensor data. Manual digitization and subsequent conversion may introduce error into the database, which may not be acceptable (Lunetta et al., 1991).

Kennedy et al. (2009) point out that this particular change detection method cannot be extended indefinitely. Land-cover classes in the original map are continually degraded by change, reducing their spectral fidelity over time and requiring eventual creation of a new land-cover map.



Advantages
* May reduce change detection errors (omission and commission)
* Provides "from–to" change class information

Disadvantages
* Requires a number of steps
* Dependent on quality of "change/no-change" binary mask

FIGURE 12–37 Diagram of change detection using a binary change mask applied to Date 2 (imagery courtesy of Richland County GIS Department).

Change Detection Using a Binary Change Mask Applied to Date 2

This method of change detection is very effective. First, the analyst selects the base image referred to as Date 1 at time n. Date 2 may be an earlier image ($n - 1$) or a later image ($n + 1$). A traditional classification of Date 1 is performed using rectified remote sensor data. Next, one of the bands (e.g., band 2 in Figure 12-37) from both dates of imagery is placed in a new dataset. The two-band dataset is then analyzed using various image algebra change detection functions (e.g., band ratio, image differencing) to produce a new change image file. The analyst selects a threshold value to identify areas of "change" and "no change" in the new image, as discussed in the section on image algebra change detection. The change image is then recoded into a binary mask file consisting of areas that have changed

between the two dates. Great care must be exercised when creating the "change/no-change" binary mask (Jensen et al., 1993a).The change mask is then overlaid onto Date 2 of the analysis and only those pixels that were detected as having changed are classified in the Date 2 imagery. A traditional post-classification comparison can then be applied using the Date 1 and Date 2 classifications to yield more detailed "from–to" change information.

This method may reduce change detection errors (omission and commission) and it provides detailed "from–to" change class information. The technique reduces effort by allowing analysts to focus on the small amount of area that has changed between dates. In most regional projects, the amount of actual change over a 1- to 5-year period is probably no greater than 10% of the total area. The method is complex, requiring a number of steps, and the final outcome is dependent on the quality of the "change/no–change" binary mask used in the analysis.

USGS scientists used a version of this technique to update the 2001 National Land Cover Database (NLCD) to 2006 using Landsat TM imagery (Xian et al., 2009). They used the 2001 NLCD data as the Date 1 classification map. They processed 2001 and 2006 Landsat imagery to identify pixels that changed between 2001 and 2006. They then classified only the pixels that changed to create the new 2006 land-cover dataset. Pixels that did not change in the 2006 land-cover map retained the land cover present in the 2001 NLCD.

Chi-Square Transformation Change Detection

Ridd and Liu (1998) introduced a chi-square transformation change detection algorithm. It will work on any type of imagery but let us for the moment apply it to six bands of Landsat TM data obtained on two dates. The chi-square transformation is:

$$Y_{pixel} = (X - M)^T \Sigma^{-1} (X - M) \tag{12.13}$$

where Y_{pixel} is the digital value of the pixel in the output change image, X is the vector of the difference of the six digital values between the two dates for each pixel, M is the vector of the mean residuals of each band for the entire image, T is the transverse of the matrix, and Σ^{-1} is the inverse covariance matrix of the six bands between the two dates. The usefulness of the transformation in this context rests on the fact that Y is distributed as a chi-square random variable with p degrees of freedom where p is the number of bands. $Y = 0$ represents a pixel of no change. The user creates the output change image and highlights pixels with varying amounts of Y.

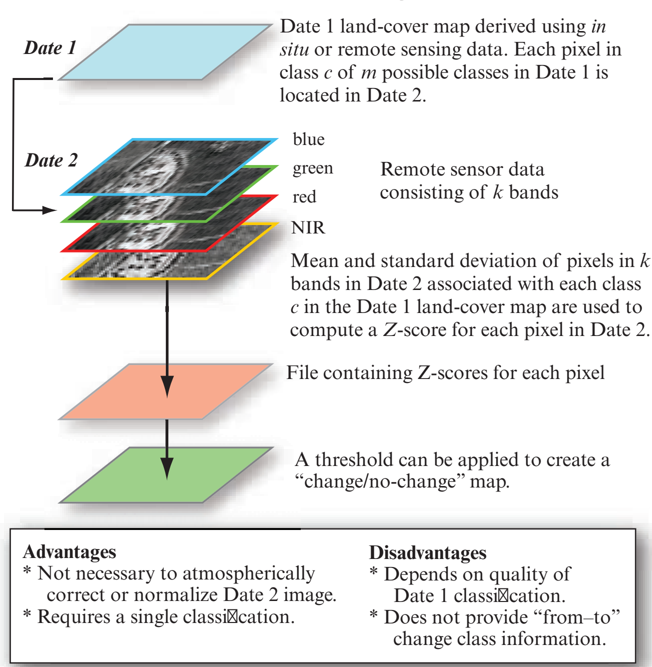


FIGURE 12–38 Diagram of cross-correlation change detection (imagery courtesy of Richland County GIS Department).

Cross-Correlation Change Detection

Cross-correlation change detection makes use of an existing Date 1 digital land-cover map and a Date 2 unclassified multispectral dataset (Koeln and Bissonnette, 2000). The Date 2 multispectral dataset does not need to be atmospherically corrected or converted to percent reflectance; the original brightness values are sufficient. Several passes through the datasets are required to implement cross-correlation change detection. First, every pixel in the Date 1 land-cover map associated with a particular class c (e.g., forest) out of m possible classes is located in the Date 2 multispectral dataset (Figure 12-38). The mean (μ_{ck}) and standard deviation (σ_{ck}) of all the brightness values in each band k in the Date 2 multispectral dataset associated with class c (e.g., forest) in the Date 1 land-cover map are computed. Next, every pixel (BV_{ijk}) in the Date 2 scene associated with class c is compared with the mean (μ_{ck}) and divided by the standard deviation (σ_{ck}). This value is summed and squared over all k bands. This is performed for each class. The result is a Z-score associated with each pixel in the scene:

$$Z_{ijc} = \sum_{k=1}^{n} \left(\frac{BV_{ijk} - \mu_{c_k}}{\sigma_{c_k}} \right)^2 \tag{12.14}$$

where

Z_{ijc} is the Z-score for a pixel at location i,j in the Date 2 multispectral dataset associated with a particular class c found in the Date 1 land-cover map;

c is the Date 1 land-cover class under investigation;

n is the number of bands in the Date 2 multispectral image;

k is the band number in the Date 2 multispectral image;

BV_{ijk} is the brightness value of a pixel (or reflectance) at location i,j in band k of the Date 2 multispectral dataset associated with a particular class found in the Date 1 land-cover map;

μ_{ck} is the mean of all pixel brightness values found in band k of the Date 2 multispectral dataset associated with a particular class c in the Date 1 land-cover map;

σ_{ck} is the standard deviation of all pixel brightness values found in band k of the Date 2 multispectral dataset associated with a particular class c in the Date 1 land-cover map.

The mean (μ_{ck}) and standard deviation (σ_{ck}) values derived from the cross-correlation of a Date 1 four-class land-cover map with a Date 2 three-band multispectral dataset would be stored in a table like Table 12-5 and used in Equation 12.14.

The Z-statistic describes how close a pixel's response is to the expected spectral response of its corresponding class value in the land-cover map (Civco et al., 2002). In the output file, the greater the Z-score of an individual pixel, the greater the probability that its land cover has changed from Date 1 to Date 2. If desired, the image analyst can examine the Z-score image file and select a threshold that can be used to identify all pixels in the scene that have changed from Date 1 to Date 2. This information can be used to prepare a "change/no-change" map of the area.

Advantages associated with cross-correlation change detection include the fact that it is not necessary to perform an atmospheric correction on any dataset. It also eliminates the problems associated with phenological differences between dates of imagery. Unfortunately, this change detection method is heavily dependent upon the accuracy of the Date 1 land-cover classification. If it has serious error, then the cross-correlation between Date 1 and Date 2 will contain error. Every

TABLE 12–5 Hypothetical mean and standard deviations associated with a cross-correlation change detection analysis of a Date 1 land-cover map with four classes and a Date 2 multispectral image that contains three bands.

Land-Cover Class	Band 1		Band 2		Band 3	
1	$\mu_{1,1}$	$\sigma_{1,1}$	$\mu_{1,2}$	$\sigma_{1,2}$	$\mu_{1,3}$	$\sigma_{1,3}$
2	$\mu_{2,1}$	$\sigma_{2,1}$	$\mu_{2,2}$	$\sigma_{2,2}$	$\mu_{2,3}$	$\sigma_{2,3}$
3	$\mu_{3,1}$	$\sigma_{3,1}$	$\mu_{3,2}$	$\sigma_{3,2}$	$\mu_{3,3}$	$\sigma_{3,3}$
4	$\mu_{4,1}$	$\sigma_{4,1}$	$\mu_{4,2}$	$\sigma_{4,2}$	$\mu_{4,3}$	$\sigma_{4,3}$

pixel in the Date 1 land-cover map must be assigned to a class. The method does not produce any "from–to" change detection information.

Visual On-Screen Change Detection and Digitization

A great amount of high-resolution remote sensor data is now available from satellite (e.g., IKONOS, GeoEye-1, WorldView-2, Pleiades) (Belward and Skoien, 2014). In addition, digital aerial photography is collected on an ongoing basis for many city, county, region, and state applications. Digital aerial photography is almost always collected rapidly after a major natural or human-made disaster. It is not always necessary or practical to use the aforementioned digital image processing change detection techniques. Rather, less complex techniques can be used. For example, it is possible to display two registered images on a computer screen and perform manual on-screen digitization to identify change polygons. This can be done using a digital image processing system or a GIS.

Hurricane Hugo Example

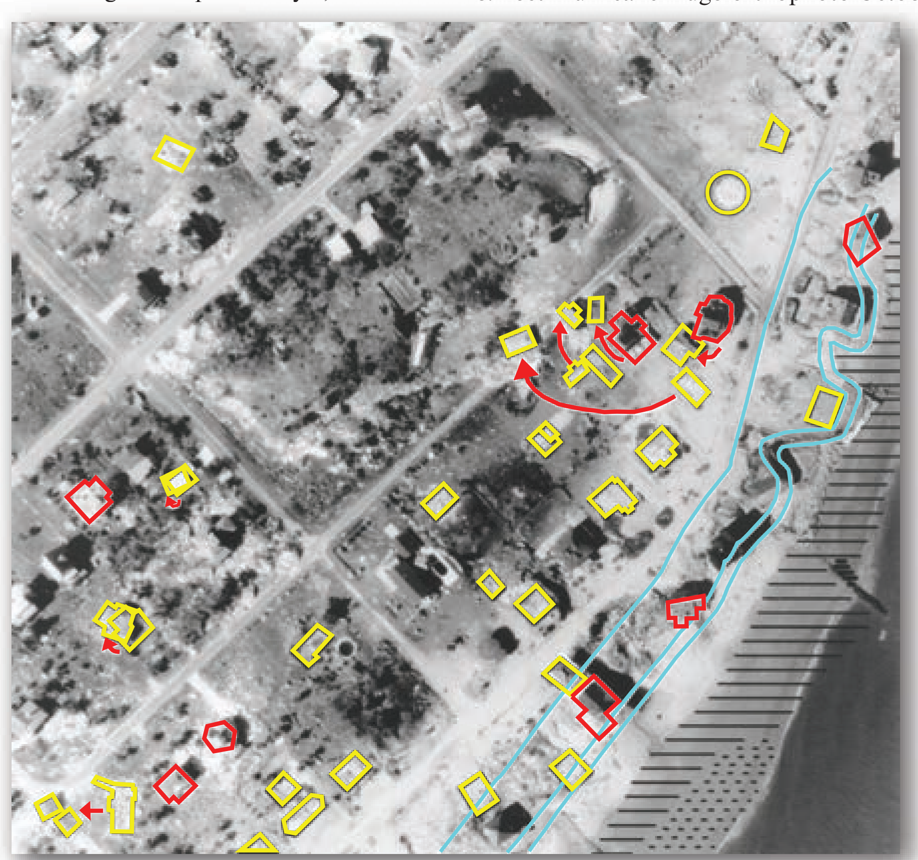
For example, consider the pre- and post-Hurricane Hugo black-and-white aerial photography shown in Figure 12-39ab. You would assume that it would be straightforward to use one of the change detection techniques (e.g., image differencing, change vector analysis) to identify the houses that were destroyed when Hurricane Hugo made landfall on September 21, 1989. Unfortunately, due to the tremendous amount of sand and debris washed ashore during the hurricane, such digital techniques were of little value. The most useful method of identifying the important changes was to simply photointerpret the two images on the screen and then digitize the important information onto the post-Hurricane Hugo image (Figure 12-39c). The photointerpretation was performed using the fundamental elements of image interpretation such as size, shape, shadow, tone, color, texture, site and association (Jensen, 2007). Homes that were completely destroyed

Visual On-Screen Change Detection of the Impact of Hurricane Hugo on Sullivan's Island, SC



a. Pre-Hurricane Hugo orthophoto July 1, 1988.

b. Post-Hurricane Hugo orthophoto October 5, 1989.



c. Change information annotated on the October 5, 1989, orthophoto.

FIGURE 12–39 a) Panchromatic orthophoto of Sullivan's Island, SC, obtained on July 1, 1988, prior to Hurricane Hugo. The data were rectified to State Plane Coordinates and resampled to 0.3×0.3 m spatial resolution. b) Panchromatic orthophoto obtained on October 5, 1989, after Hurricane Hugo. c) Change information overlaid on October 5, 1989, post-Hurricane Hugo aerial photograph, Sullivan's Island, SC. Completely destroyed houses are outlined in red. Partially destroyed houses are outlined in yellow. A red arrow indicates the direction of houses removed from their foundations. Three beachfront management setback lines are shown in cyan (base, 20 year, 40 year). Areas of beach erosion are depicted as black lines. Areas of beach accretion caused by Hurricane Hugo are shown as dashed black lines.

are outlined in red, partially destroyed homes are outlined in yellow, and the movement of homes off their foundations is shown using red lines with arrowheads (Figure 12-39c).

Hurricane Katrina Example

The same logic was applied to identify important change associated with Hurricane Katrina, which made landfall on August 29, 2005 (Figure 12-40a). Low oblique and vertical aerial photography of the effects of Hurricane Katrina in Gulfport, MS, are shown in Figure 12-40bc. Hurricane Katrina storm surge did great damage to coastal geomorphology, vegetation, and especially human-made structures. A USGS high spatial resolution (1.2 × 1.2 ft.) color vertical aerial photograph of the Grand Casino in Gulfport, MS, is shown in Figure 12-40b. Note that approximately one-fourth of the Grand Casino was broken off by storm surge and moved onto Highway 90. Also note how numerous marine cargo containers are now located inland. The low oblique aerial photography makes it easy to determine what change has taken place with the large section of the Grand Casino located in the center of Highway 90 (Figure 12-40c). Also note the debris piled up against the hotel in the background.

Aral Sea Example

Sometimes it is useful to simply visually examine multiple dates of remotely sensed imagery to appreciate *processes* at work. For example, consider the change in the Aral Sea in Kazakhstan from 1973 to 2000. The shoreline for a portion of the Aral Sea recorded by the Landsat MSS in 1973 and 1987 and the Landsat ETM$^+$ in 2000 is shown in Figure 12-41.

The Aral Sea is actually a lake, a body of fresh water. Unfortunately, more than 60 percent of the lake has disappeared in the last 30 years. Farmers and state offices in Uzbekistan, Kazakhstan, and Central Asian states began diverting river water to the lake in the 1960s to irrigate cotton fields and rice paddies. In 1965, the Aral Sea received about 50 cubic kilometers of fresh water per year—a number that fell to zero by the early 1980s. Concentrations of salts and minerals began to rise in the lake. The change in chemistry led to alterations in the lake's ecology, causing precipitous drops in the fish population. The commercial fishing industry employed 60,000 people in the early 1960s. By 1977, the fish harvest was reduced by 75 percent, and by the early 1980s the commercial fishing industry was gone.

The shrinking Aral Sea has also had a noticeable effect on the region's climate. The growing season is now shorter, causing many farmers to switch from cotton to rice, which requires even more diverted water. A secondary effect of the reduction in the Aral Sea's overall size is the rapid exposure of the lake bed. Strong winds that blow across this part of Asia routinely pick up and deposit tens of thousands of tons of now-exposed soil every year. This process has caused a reduction in air quality for nearby residents and affected crop yields due to the salt-laden particles falling on arable land.

Environmental experts agree that the current situation cannot be sustained. Yet, driven by poverty and their dependence upon exports, officials in the region have failed to take any preventive action, and the Aral Sea continues to shrink (NASA Aral Sea, 2014).

National Land Use/Cover Database of China Example

Visual on-screen digitization change detection was recently used to identify change in the National Land Use/Cover Database of China at 1:100,000 scale using primarily Landsat TM imagery (Zhang et al.,2014).



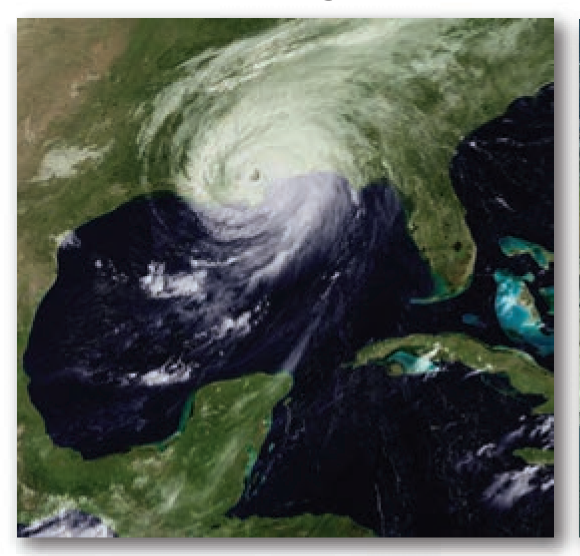
Atmospheric Correction for Change Detection

Now that many of the most widely adopted change detection algorithms have been identified, it is useful to provide some general guidelines about when it is necessary to atmospherically correct the individual dates of imagery used in the change detection process.

When Atmospheric Correction Is Necessary

Atmospheric correction of multiple-date remote sensor data is required when the individual date images used in the change detection algorithm are based on linear transformations of the data, e.g., a normalized difference vegetation index image is produced for Date 1 and Date 2. The additive effects of the atmosphere on each date contaminate the NDVI values and the modification is not linear (Song et al., 2001). Contributions from the atmosphere to NDVI values are significant and can amount to 50% or more over thin or broken vegetation cover (McDonald et al., 1998; Song et al., 2001). Similarly, the imagery should be atmospherically corrected if the change detection is based on multiple-date red/near-infrared ratioed images (e.g., Landsat TM 4/TM 3). This suggests that it may be necessary to normalize or atmospherically correct the multiple-date imagery used to compute the linearly transformed data (e.g., NDVI) when the goal is to identify *biophysical* change characteristics through time rather than just land-cover change through time (Yuan and Elvidge, 1998; Song et al., 2001; Du et al., 2002).

Visual On-Screen Change Detection of Hurricane Katrina Damage at Gulfport, MS




a. GOES-⌧⌧ August ⌧⌧, ⌧⌧⌧⌧, visible band image overlaid on MODIS true-color background.



c. Low-oblique photograph of a part of the Grand Casino moved by storm surge.



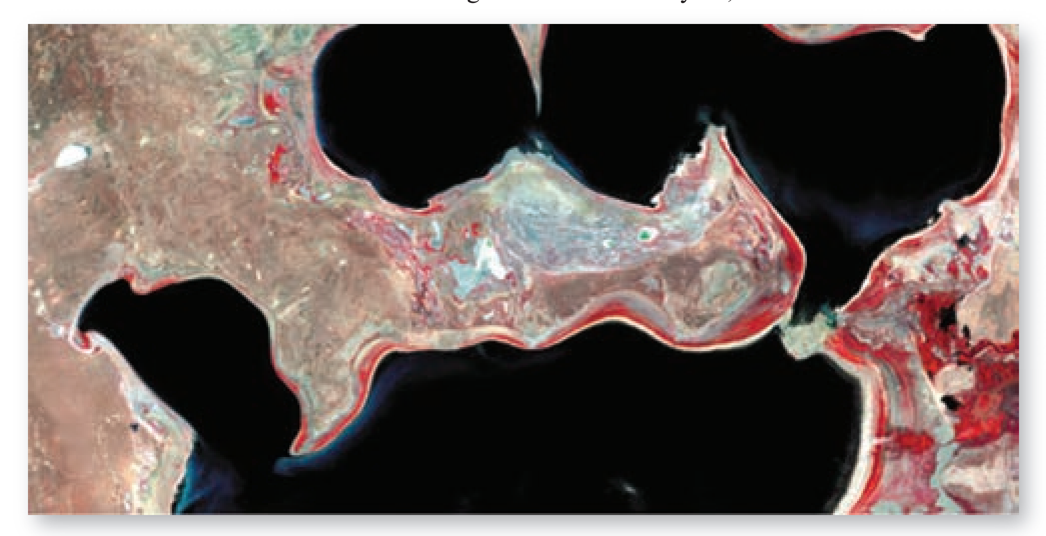
b. Vertical aerial photograph of the Grand Casino and surrounding area obtained at a spatial resolution of ⌧7 × ⌧7 cm (⌧.⌧ × ⌧.⌧ ft).

FIGURE 12–40 a) GOES-12 image of Hurricane Katrina making landfall on August 29, 2005, overlaid on MODIS true-color imagery (image courtesy of USGS Coastal and Marine Geology Program). b) Vertical high spatial resolution aerial photograph of the Grand Casino in Gulfport, MS, on August 29, 2005 (image courtesy of USGS Coastal and Marine Geology Program). c) Low oblique aerial photograph documenting the location of a part of the Grand Casino now lying on Highway 90 (photography courtesy of NOAA Remote Sensing Division).

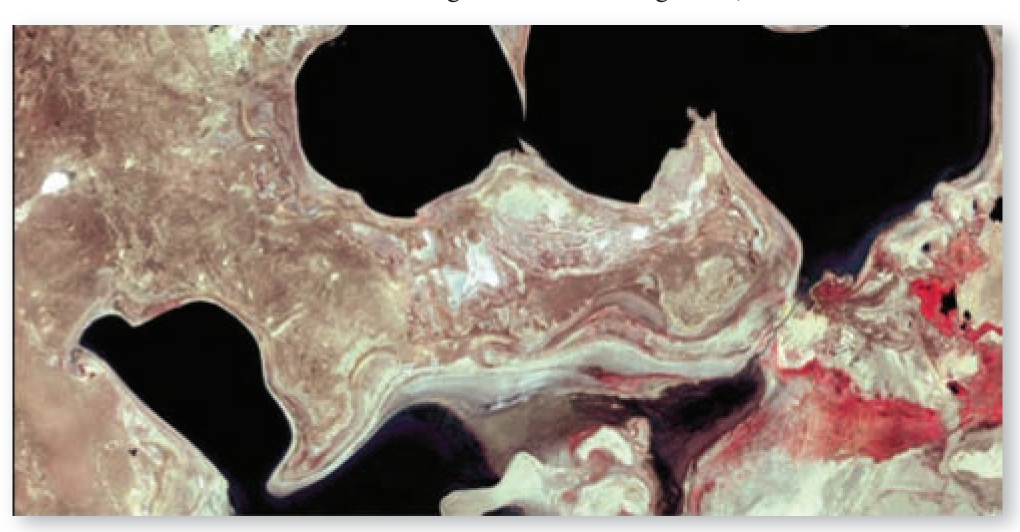
Drawdown of the Aral Sea, Kazakhstan, from 1973 to 2000



a. Landsat MSS image obtained on May [illegible], [illegible]7[illegible].



b. Landsat MSS image obtained on August [illegible], [illegible]87.



c. Landsat 7 ETM$^+$ image obtained on July [illegible], [illegible].

FIGURE 12–41 A portion of the Aral Sea shoreline recorded by the Landsat Multispectral Scanner (MSS) and Landsat Enhanced Thematic Mapper Plus (ETM$^+$) in 1973, 1987, and 2000 (imagery courtesy of NASA Earth Observatory).

A change/no-change map produced using image differencing logic and atmospherically corrected remote sensor data normally looks different from a change/no-change map produced using image differencing logic and non-atmospherically corrected data if the threshold boundaries are held constant in the change image histograms. However, if the analyst selects the appropriate thresholds in the two tails of the change detection image histogram, it doesn't really matter whether the change detection map was produced using atmospherically corrected or non-atmospherically corrected data. But, if the analyst desires that all stable classes in the change image have a value of 0 in the change histogram then it is useful to normalize one image to another or atmospherically correct both images prior to performing image differencing.

Obtaining quality training data is very expensive and time-consuming because it usually involves people and fieldwork. Therefore, it will become increasingly important to be able to extend training data through both time and space. In other words, training data extracted from a Date 1 image should be able to be extended to a Date 2 image of the same geographic area (signature extension through time) or perhaps even to a Date 1 or Date 2 image of a neighboring geographic area (signature extension through space). Extending training data through space and time will require that each image evaluated be atmospherically corrected to surface reflectance whenever possible using one of the techniques described in Chapter 6. Nevertheless, it is not always necessary to correct remote sensor data when classifying individual dates of imagery or performing change detection.

When Atmospheric Correction Is Unnecessary

A number of studies have documented that it is unnecessary to correct for atmospheric effects prior to image classification if the spectral signatures characterizing the desired classes are derived from the image to be classified. This is because atmospherically correcting a single date of imagery is often equivalent to subtracting a constant from all pixels in a spectral band. This action simply translates the origins in multidimensional feature space. The class means may change, but the variance–covariance matrix remains the same irrespective of atmospheric correction. In other words, atmospheric correction is unnecessary as long as the training data and the data to be classified are in the same relative scale (corrected or uncorrected) (Song et al., 2001). This suggests that it is not necessary to atmospherically correct Landsat TM data obtained on Date 1 if it is going to be subjected to a maximum likelihood classification algorithm and all the training data are derived from the Date 1 imagery. The same holds true when a Date 2 image is classified using training data extracted from the Date 2 image. Change between the Date 1 and Date 2 classification maps derived from the individual dates of imagery (corrected or uncorrected) can easily be compared in a post-classification comparison.

Atmospheric correction is also unnecessary when change detection is based on classification of multiple-date composite imagery in which the multiple dates of remotely sensed images are rectified and placed in a single dataset and then classified as if it were a single image (e.g., multiple-date principal components change detection). Only when training data from one time and/or place are applied in another time and/or place is atmospheric correction necessary for image classification and many change detection algorithms. The use of relative radiometric correction to normalize intensities of bands of multitemporal imagery to a reference scene is sufficient for many change detection applications (Song et al., 2001; Coppin et al., 2004; Chen et al., 2012).



Summary

A one-time inventory of natural resources is often of limited value. A time series of images and the detection of change provides significant information on *the resources at risk* and may be used in certain instances to identify *the agents of change*. Change information is becoming increasingly important in local, regional, and global environmental monitoring (Woodcock et al., 2001; Olofsson et al., 2014). This chapter identifies the remote sensor system and environmental variables that should be considered whenever a remote sensing change detection project is initiated. Several useful change detection algorithms were reviewed. Scientists are encouraged to carefully review and understand these principles so that accurate change detection can take place.



References

Ahlqvist, O., 2008, "Extending Post-classification Change Detection using Semantic Similarity Metrics to Overcome Class Heterogeneity: A Study of 1992 and 2001 U.S. National Land Cover Database Changes," *Remote Sensing of Environment*, 112:1226–1241.

Almutairi, A., and T. A. Warner, 2010, "Change Detection Accuracy and Image Properties: A Study Using Simulated Data," *Remote Sensing*, 2:1508–1529.

Arzandeh, S., and J. Wang, 2003, "Monitoring the Change of *Phragmites* Distribution Using Satellite Data," *Canadian Journal of Remote Sensing*, 29(1):24–35.

Belward, A., and J. Skoien, 2014, "Who Launched What, When and Why? Trends in Global Land-cover Observation Capacity from Civilian Earth Observation Satellites," *ISPRS Journal of Photogrammetry & Remote Sensing*, http://dx.doi.org/10.1016/j.isprsjprs.2014.03.009.

Bruzzone, L., and D. F. Prieto, 2000, "Automatic Analysis of the Difference Image for Unsupervised Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1171–1182.

Carvalho, O. A., Guimaraes R. F., Gillespie, A. R., Silva, N. C., and R. A. T. Gomes, 2011, "A New Approach to Change Vector Analysis Using Distance and Similarity Measures," *Remote Sensing*, 3:2473–2493.

Celik, T., 2009, "Unsupervised Change Detection in Satellite Images Using Principal Component Analysis and k-Means Clustering," *IEEE Geoscience & Remote Sensing Letters*, 6(4):772–776.

Chen, G., Hay, G. J., Carvalho, L. M. T., and M. A. Wulder, 2012, "Object-based Change Detection," *International Journal of Remote Sensing*, 33(14):4434–4457.

Chen, J., Gong, P., He, C., Pu, R., and P. Shi, 2003, "Land-Use/Land-Cover Change Detection Using Improved Change-Vector Analysis," *Photogrammetric Engineering & Remote Sensing* 69(4):369–379.

Civco, D. L., Hurd, J. D., Wilson, E. H., Song, M., and Z. Zhang, 2002, "A Comparison of Land Use and Land Cover Change Detection Methods," *Proceedings, ASPRS-ACSM Annual Conference and FIG XXII Congress,* Bethesda: American Society for Photogrammetry & Remote Sensing, 10 p, CD.

Collins, J. B., and C. E. Woodcock, 1996, "An Assessment of Several Linear Change Detection Techniques for Mapping Forest Mortality Using Multitemporal Landsat TM Data," *Remote Sensing of Environment*, 56:66–77.

Congalton, R. G., and K. Green, 2009, *Assessing the Accuracy of Remotely Sensed Data—Principles and Practices*, (2nd Ed.), Boca Raton: CRC Press, 183.

Coppin, P., Jonckheere, I., Nackaerts, K., and B. Muys, 2004, "Digital Change Detection Methods in Ecosystem Monitoring: A Review," *International Journal of Remote Sensing*, 25(9):1565–1596.

Coulter, L., Hope, A., Stow, D., Lippitt, C., and S. Lathrop, 2011, "Time-space Radiometric Normalization of TM/ETM+ Images for Land Cover Change Detection," *International Journal of Remote Sensing*, 32(22):7539–7556.

Crippin, R. E., 1988, "The Dangers of Underestimating the Importance of Data Adjustments in Band Ratioing," *International Journal of Remote Sensing*, 9(4):767–776.

Deng, J. S., Wang, K., Deng, Y. H., and G. J. Qi, 2008, "PCA-based Land-use Change Detection and Analysis using Multitemporal Multisensor Satellite Data," *International Journal of Remote Sensing*, 29(16):4823–4838.

Dobson, J. E., Bright, E. A., Ferguson, R. L., Field, D. W., Wood, L. L., Haddad, K. D., Iredale, H., Jensen, J. R., Klemas, V. V., Orth, R. J., and J. P. Thomas, 1995, *NOAA Coastal Change Analysis Program (C-CAP)—Guidance for Regional Implementation*, Washington: NOAA, 92 p.

Du, Y., Teillet, P. M., and J. Cihlar, 2002, "Radiometric Normalization of Multitemporal High-resolution Satellite Images with Quality Control for Land Cover Change Detection," *Remote Sensing of Environment*, 82:123–134.

Dykstra, J., 2012, "Comparison of Landsat and RapidEye Data for Change Monitoring Applications," presented at the *11th Annual Joint Agency Commercial Imagery Evaluation (JACIE) Workshop*, Fairfax, VA, April 17–19, 32 p.

Elaksher, A. F., Bethel, J. S., and E. M. Mikhail, 2003, "Roof Boundary Extraction Using Multiple Images," *Photogrammetric Record*, 18(101):27–40.

Esri ChangeMatters, 2014, *Change Matters*, Redlands: Esri, Inc., http://changematters.esri.com/compare.

Exelis, 2013, *ENVI EX: Image Difference Change Detection*, Boulder: Exelis Visual Information Solutions, 9 p.

Foley, J., DeFries, R., Asner, G., Barford, C., Bonan, G., Carpenter, S., Chapin, F., Coe, M. Daily, C., Gibbs, H., Helkowski, J., Holloway, T., Howard, E., Kucharik, C., Monfreda, C., Patz, J., Prentice, I., Ramankutty, N., and P. Snyder, 2005, "Global Consequences of Land Use," *Science*, July 5, 309:570–574.

Foody, G. M., 2001, "Monitoring Magnitude of Land-Cover Change Around the Southern Limits of the Sahara," *Photogrammetric Engineering & Remote Sensing*, 67(7):841–847.

Franklin, S. E., Lavigne, M. B., Wulder, M. A., and T. M. McCaffrey, 2002, "Large-area Forest Structure Change Detection: An Example," *Canadian Journal of Remote Sensing*, 28(4):588–592.

Fry, J. A., Xian, G., Jin, S., Dewitz, J. A., Homer, C. G., Yang, L., Barnes, C. A., Herold, N. D., and J. D. Wickham, 2011, "National Land Cover Database for the Conterminous United States," *Photogrammetric Engineering & Remote Sensing*, 77(9):859–863.

Gartner, P., Forster, M., Kurban, A., and B. Kleinschmit, 2014, "Object Based Change Detection of Central Asian Tugai Vegetation with Very High Spatial Resolution Satellite Imagery," *International Journal of Applied Earth Observation and Geoinformation*, 31:110–121.

Gladstone, C. S., Gardiner, A., and D. Holland, 2012, "A Semi-Automatic Method for Detecting Changes to Ordinance Survey Topographic Data in Rural Environments," *Proceedings of the 4th GEOBIA Conference*, May 7–9, Rio de Janeiro, Brazil, 396–401.

Green, K., 2011, "Change Matters," *Photogrammetric Engineering & Remote Sensing*, 77(4):305–309.

Griffiths, P., Hostert, P., Gruebner, O., and S. van der Linden, 2010, "Mapping Megacity Growth with Multi-sensor Data," *Remote Sensing of Environment*, 114:426–439.

Hofmann, P., Strobl, J., Blaschke, T., and H. J. Kux, 2008, "Detecting Informal Settlements from QuickBird Data in Rio de Janeiro Using an Object-based Approach," in *Object-based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*, Blaschke, T., Lang, S. and G. J. Hay, Eds., Springer: Berlin, 531–554.

Huang, X., and M. A. Friedl, 2014, "Distance Metric-based Forest Cover Change Detection using MODIS Time Series," *International Journal of Applied Earth Observation and Geoinformation*, 29:78–92.

Hurskainen, P., and P. Pellikka, 2004, "Change Detection of Informal Settlements Using Multi-temporal Aerial Photographs—The Case of Voi, SE-Kenya," in *Proceedings of the AARSE Conference*, Nairobi, Kenya, 18–21 October.

Im, J., 2006, *A Remote Sensing Change Detection System Based on Neighborhood/Object Correlation Image Analysis, Expert Systems, and an Automated Calibration Model*, Ph.D. dissertation, Dept. of Geography, Univ. of South Carolina, Columbia, SC.

Im, J., and J. R. Jensen, 2005, "Change Detection Using Correlation Analysis and Decision Tree Classification," *Remote Sensing of Environment*, 99:326–340.

Im, J., Jensen, J. R., and J. A. Tullis, 2008, "Object-based Change Detection Using Correlation Image Analysis and Image Segmentation," *International Journal of Remote Sensing*, 29(2):399–423.

Im, J., Jensen, J. R., Jensen, R. R., Gladden, J., Waugh, J., and M. Serrato, 2012, "Vegetation Cover Analysis of Hazardous Waste Sites in Utah and Arizona using Hyperspectral Remote Sensing," *Remote Sensing*, (4):327–353.

Im, J., Rhee, J., and J. R. Jensen, 2009, "Enhancing Binary Change Detection Performance Using a Moving Threshold Window (MTW) Approach," *Photogrammetric Engineering & Remote Sensing*, 75(8):951–962.

Im, J., Rhee, J., Jensen, J. R., and M. E. Hodgson, 2007, "An Automated Binary Change Detection Model Using a Calibration Approach," *Remote Sensing of Environment*, 106:89–105.

Im, J., Rhee, J., Jensen, J. R., and M. E. Hodgson, 2008, "Optimizing the Binary Discriminant Function in Change Detection Applications," *Remote Sensing of Environment*, 112:2761–2776.

Im, J., Lu, Z., and J. R. Jensen, 2011, "A Genetic Algorithm Approach to Moving Threshold Optimization for Binary Change Detection," *Photogrammetric Engineering & Remote Sensing*, 77(2):167–180.

Jensen, J. R., 1981, "Urban Change Detection Mapping Using Landsat Digital Data," *The American Cartographer*, 8(2):127–147.

Jensen, J. R., 2007, *Remote Sensing of the Environment: An Earth Resource Perspective*, (2nd Ed.), Boston: Pearson, 592 p.

Jensen, J. R., and J. Im., 2007, "Remote Sensing Change Detection in Urban Environments," in R.R. Jensen, J. D. Gatrell and D. D. McLean (Eds.), *Geo-Spatial Technologies in Urban Environments Policy, Practice, and Pixels*, (2nd Ed.), Berlin: Springer-Verlag, 7–32.

Jensen, J. R., Cowen, D., Narumalani, S., Althausen, J., and O. Weatherbee, 1993a, "Evaluation of CoastWatch Change Detection Protocol in South Carolina," *Photogrammetric Engineering & Remote Sensing*, 59(6):1039–1046.

Jensen, J. R., Guptill, S., and D. Cowen, 2012, *Change Detection Technology Evaluation*, Bethesda: U.S. Bureau of the Census, Task T007, FY2012 Report, 232 p.

Jensen, J. R., Huang, X., and H. E. Mackey, 1997, "Remote Sensing of Successional Changes in Wetland Vegetation as Monitored During a Four-Year Drawdown of a Former Cooling Lake," *Applied Geographic Studies*, 1:31–44.

Jensen, J. R., Im, J., Hardin, P., and R. R. Jensen, 2009, "Chapter 19: Image Classification," in *Remote Sensing*, Warner, T., Nellis, D. and G. Foody, Eds., Los Angeles: Sage Publications, 269–296.

Jensen, J. R., Narumalani, S., Weatherbee, O., and H. E. Mackey, 1993b, "Measurement of Seasonal and Yearly Cattail and Waterlily Changes Using Multidate SPOT Panchromatic Data," *Photogrammetric Engineering & Remote Sensing*, 59(4):519–525.

Jensen, J. R., Rutchey, K., Koch, M., and S. Narumalani, 1995, "Inland Wetland Change Detection in the Everglades Water Conservation Area 2A Using a Time Series of Normalized Remotely Sensed Data," *Photogrammetric Engineering & Remote Sensing*, 61(2):199–209.

Kennedy, R. E., Townsend, P. A., Gross, J. E., Cohen, W. B., Bolstad, P., Wang, Y. Q., and P. Adams, 2009, "Remote Sensing Change Detection Tools for Natural Resource Managers: Understanding Concepts and Tradeoffs in the Design of Landscape Monitoring Projects," *Remote Sensing of Environment*, 113:1382–1396.

Kit, O., and M. Ludeke, 2013, "Automated Detection of Slum Area Change in Hyderabad, India using Multitemporal Satellite Imagery," *ISPRS Journal of Photogrammetry & Remote Sensing*, 83:130–137.

Klemas, V., 2011, "Remote Sensing of Wetlands: Case Studies Comparing Practical Techniques," *Journal of Coastal Research*, 27(3):418–427.

Koeln, G., and J. Bissonnette, 2000, "Cross-correlation Analysis: Mapping Land Cover Change with a Historic Land Cover Database and a Recent, Single-date Multispectral Image," *Proceedings, American Society for Photogrammetry & Remote Sensing,* Bethesda: ASP&RS, 8 p., CD.

Kontoes, C. C., 2013, “Operational Land Cover Change Detection using Change Vector Analysis,” *International Journal of Remote Sensing*, 29(16):4757–4779.

Leprince, S., Barbot, S., Ayoub, F., and J. Avouac, 2007, “Automatic and Precise Orthorectification, Coregistration, and Subpixel Correlation of Satellite Images, Application to Ground Deformation Measurements,” *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1529–1588.

Lu, D., Mausel, P., Brondizio, E., and E. Moran, 2004, “Change Detection Techniques,” *International Journal of Remote Sensing*, 25(12):2365–2401.

Lunetta, R. S., and C. Elvidge, 2000, *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*, New York: Taylor & Francis, 340 p.

Lunetta, R. S., Congalton, R. G., Fenstermaker, L. K., Jensen, J. R., McGwire, K. C., and L. R. Tinney, 1991, “Remote Sensing and Geographic Information System Data Integration: Error Sources and Research Issues,” *Photogrammetric Engineering & Remote Sensing,* 57(6):677–687.

Lunetta, R. S., Knight, J. F., Ediriwickrema, J., Lyon, J. G., and L. D. Worthy, 2006, “Land-cover Change Detection using Multi-temporal MODIS NDVI Data,” *Remote Sensing of Environment*, 105:142–154.

Malila, W. A., 1980, “Change Vector Analysis: An Approach for Detecting Forest Changes with Landsat,” *Proceedings, LARS Machine Processing of Remotely Sensed Data Symposium,* W. Lafayette, IN: Laboratory for the Applications of Remote Sensing, 326–336.

Matikainen, L., Hyypp⌧, J., Ahokas, E., Markelin, L., and H. Kaartinen, 2010, “Automatic Detection of Buildings and Changes in Buildings for Updating of Maps,” *Remote Sensing*, 2010(2):1217–1248.

McDonald, A. J., Gemmell, F. M., and P. E. Lewis, 1998, “Investigation of the Utility of Spectral Vegetation Indices for Determining Information on Coniferous Forests,” *Remote Sensing of Environment*, 66:250–272.

MDA, 2014, *National Urban Change Indicator (NUCI)*, Gaithersburg: MDA Information Systems LLC, (http://www.mdaus.com and http://www.mdaus.com/Geospatial/Global-Change-Monitoring.aspx).

Michalek, J. L., Wagner, T. W., Luczkovich, J. J., and R. W. Stoffle, 1993, “Multispectral Change Vector Analysis for Monitoring Coastal Marine Environments,” *Photogrammetric Engineering & Remote Sensing*, 59(3):381–384.

Miller, R. B., Abbott, M. R., Harding, L. W., Jensen, J. R., Johannsen, C. J., Macauley, M., MacDonald, J. S., and J. S. Pearlman, 2003, *Using Remote Sensing in State and Local Government: Information for Management and Decision Making*, Washington: National Academy Press, 97 p.

Mitishita, E., Habib, A., Centeno, J., Machado, A., Lay, J., and C. Wong, 2008, “Photogrammetric and LiDAR Data Integration Using the Centroid of A Rectangular Roof as a Control Point,” *Photogrammetric Record*, 23(121):19–35.

Moeller, M. S., and T. Blaschke, 2006,”Urban Change Extraction from High Resolution Satellite Image,” *Proceedings ISPRS Technical Commission II Symposium*, Vienna, Austria, 12-14 July, 151–156.

Morisette J. T., and S. Khorram, 2000, “Accuracy Assessment Curves for Satellite-based Change Detection,” *Photogrammetric Engineering & Remote Sensing*, 66:875–880.

NASA Aral Sea, 2014, *The Shrinking Aral Sea*, Washington: NASA Earth Observatory, http://earthobservatory.nasa.gov/IOTD/view.php?id=1396.

Nielsen, A. A., and M. J. Canty, 2008, “Kernel Principal Component Analysis for Change Detection,” in L. Bruzzone (Ed.), *Image and Signal Processing for Remote Sensing* XIV, *Proceedings* of the SPIE, Volume 7109:10 p.

NRC, 2010, *New Research Directions for the National Geospatial-Intelligence Agency*, Washington: National Research Council, 59 p.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S., Woodcock, C. E., and M. A. Wulder, 2014, “Good Practices for Estimating Area and Assessing Accuracy of Land Change,” *Remote Sensing of Environment*, 148:42–57.

Patra, S., Ghosh, S., and A. Ghosh, 2013, “Histogram Thresholding for Unsupervised Change Detection of Remote Sensing Images,” *International Journal of Remote Sensing*, 32(12):6071–6089.

Pictometry International, 2014, *ChangeFindr*, http://www.eagleview.com/Products/ImageSolutionsAnalytics/PictometryAnalyticsDeployment.aspx#ChangeFinder.

Purkis, S., and V. Klemas, 2011, *Remote Sensing and Global Environmental Change*, New York: Wiley-Blackwell, 367 p.

Qin, R., 2014, “Change Detection on LOD 2 Building Models with Very High Resolution Spaceborne Stereo Imagery,” *ISPRS Journal of Photogrammetry & Remote Sensing*, 96:179–192.

Radke, R. J., Andra, S., Al-Kofahi, O., and B. Roysam, 2005, “Image Change Detection Algorithms: A Systematic Survey,” *IEEE Transactions on Image Processing*, 14(3):294–307.

Ridd, M. K., and J. Liu, 1998, “A Comparison of Four Algorithms for Change Detection in an Urban Environment,” *Remote Sensing of Environment*, 63:95–100.

Rokni, K., Ahmad, A., Solaimani, K., and S. Hazini, 2015, “A New Approach for Surface Water Change Detection: Integration of Pixel Level Image Fusion and Image Classification Techniques,” *International Journal of Applied Earth Observation and Geoinformation*, 34:226–234.

Roy, D. P., 2000, “The Impact of Misregistration Upon Composited Wide Field of View Satellite Data and Implications for Change Detection,” *IEEE Transactions on Geoscience and Remote Sensing*, 38(4):2017–2032.

Rutchey, K., and L. Velcheck, 1994, “Development of an Everglades Vegetation Map Using a SPOT Image and the

Global Positioning System," *Photogrammetric Engineering & Remote Sensing*, 60(6):767–775.

Schneider, A., 2012, "Monitoring Land Cover Change in Urban and Peri-urban Areas using Dense Time Stacks of Landsat Satellite Data and a Data Mining Approach," *Remote Sensing of Environment*, 124:689–704.

Sohn, G., and I. Dowman, 2007, "Data Fusion of High-resolution Satellite Imagery and LiDAR data for Automatic Building Extraction," *ISPRS Journal of Photogrammetry & Remote Sensing*, 62(1):43–63.

Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., and S. A. Macomber, 2001, "Classification and Change Detection Using Landsat TM Data: When and How to Correct Atmospheric Effects," *Remote Sensing of Environment*, 75:230–244.

Stow, D., Toure, S. I., Lippitt, C. D., Lippitt, C. L., and C. R. Lee, 2011, "Frequency Distribution Signatures and Classification of Within-object Pixels," *International Journal of Applied Earth Observations and Geoinformation*, 15:49-56.

Taubenbock, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., and S. Dech, 2012, "Monitoring Urbanization in Mega Cities from Space," *Remote Sensing of Environment*, 117:162–176.

Tsai, Y., Stow, D., and J. Weeks, 2011, "Comparison of Object Based Image Analysis Approaches to Mapping New Buildings in Accra, Ghana Using Multi-Temporal QuickBird Satellite Imagery," *Remote Sensing*, 3:2707–2726.

USGS, 2014, *Opening the Landsat Archive – Fact Sheet*, Washington: U. S. Geological Survey, http://pubs.usgs.gov/fs/2008/3091/pdf/fs2008-3091.pdf.

van Oort, 2007, "Interpreting the Change Detection Error Matrix," *Remote Sensing of Environment*, 108:1–8.

Virag, L. A., and J. E. Colwell, 1987, "An Improved Procedure for Analysis of Change in Thematic Mapper Image-Pairs," *Proceedings of the 21st International Symposium on Remote Sensing of Environment*, Ann Arbor: Environmental Research Institute of Michigan, 1101–1110.

Warner, T. A., Almutairi, A., and J. Y. Lee, 2009, "Remote Sensing of Land Cover Change," in Warner, T. A., Nelllis, M. D., and G. M. Foody (Eds.), *The Sage Handbook of Remote Sensing*, Los Angeles: Sage, 459–472.

Woodcock, C. E., Macomber, S. A., Pax-Lenney, M., and W. B. Cohen, 2001, "Monitoring Large Areas for Forest Change Using Landsat: Generalization Across Space, Time and Landsat Sensors," *Remote Sensing of Environment*, 78:194–203.

Xian, G., Homer, C., and J. Fry, 2009, "Updating the 2001 National Land Cover Database Land Cover Classification using Landsat Imagery Change Detection Methods," *Remote Sensing of Environment*, 113:1133–1147.

Yuan, D., and C. Elvidge, 1998, "NALC Land Cover Change Detection Pilot Study: Washington D.C. Area Experiments," *Remote Sensing of Environment*, 66:166–178.

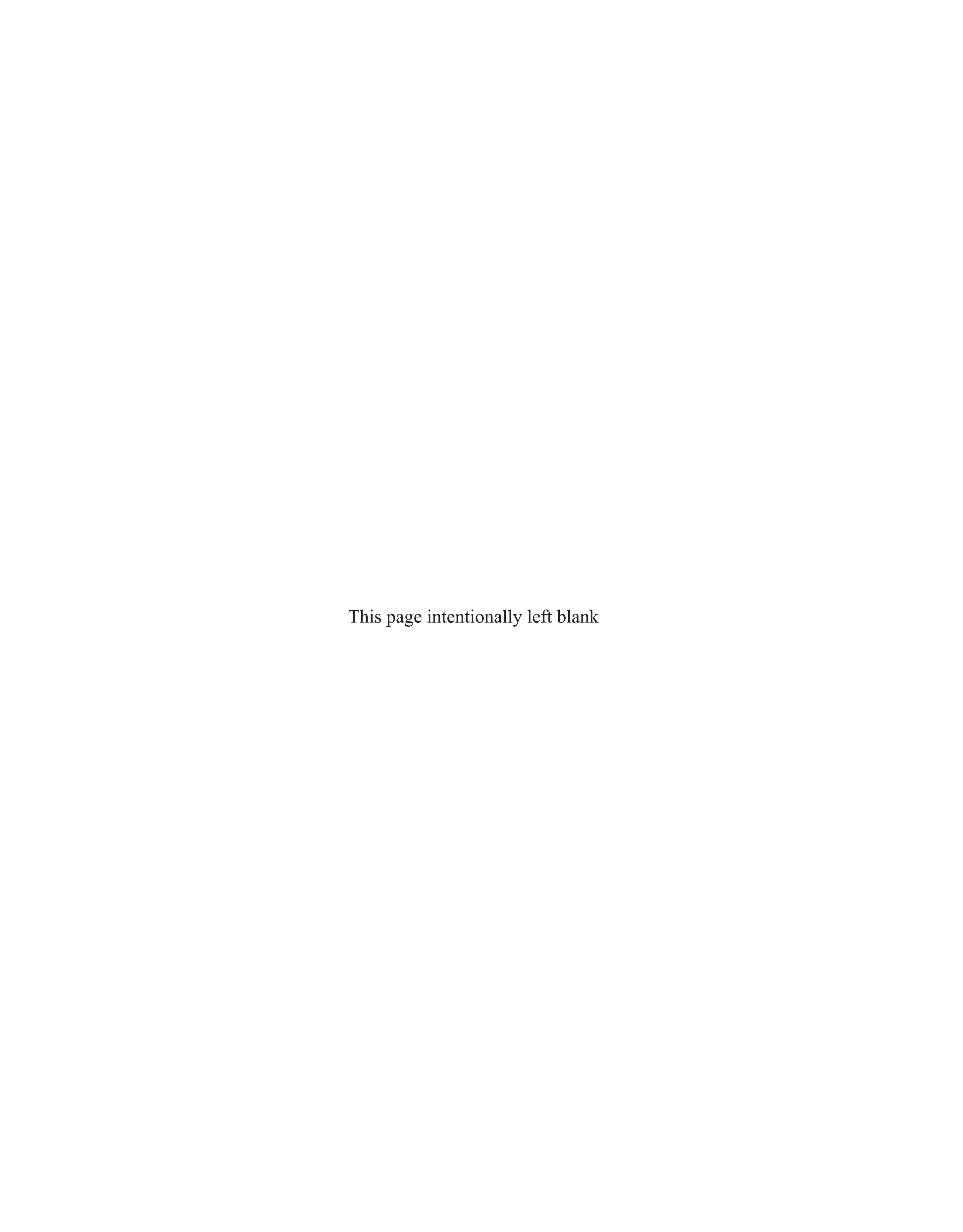
Yuan, Y., Lv, H., and X. Lu, 2015, "Semi-supervised Change Detection Method for Multi-temporal Hyperspectral Images," *Neurocomputing*, 148:363–375.

Zhan, Q., Molenaar, M., Tempfli, K., and W. Shi, 2005, "Quality Assessment for Geo-spatial Objects Derived from Remotely Sensed Data" *International Journal of Remote Sensing*, 2005(26):2953–2974.

Zhan, X., Sohlberg, R. A., Townshend, J. R. G., DiMiceli, C., Carroll, M. L., Eastman, J. C., Hansen, M. C., and R. S. DeFries, 2002, "Detection of Land Cover Changes Using MODIS 250 m Data," *Remote Sensing of Environment*, 83:336–350.

Zhang, Z., and 9 co-authors, 2014, "A 2010 Update of National Land Use/Cover Database of China at 1:100,000-scale using Medium Spatial Resolution Satellite Images," *Remote Sensing of Environment*, 149:142–152.

Zhu, Z., and C. E. Woodcock, 2014, "Continuous Change Detection and Classification of Land Cover Using All Available Landsat Data," *Remote Sensing of Environment*, 144:152–171.



This page intentionally left blank

13 Remote Sensing–Derived Thematic Map Accuracy Assessment



Information derived from remotely sensed data is becoming increasingly important for planning, ecosystem monitoring, food security and human health assessment, and police/military reconnaissance at local, regional, and global scales (NRC, 2007, 2009, 2012). The remote sensing–derived thematic information may be in the form of thematic maps or statistics derived from sampling techniques. The thematic information must be accurate because important decisions are made throughout the world using the geospatial information (Jensen and Jensen, 2013; Wickham et al., 2013).

Unfortunately, remote sensing–derived thematic information contains error. Analysts who create remote sensing–derived thematic information should recognize the sources of the error, minimize it as much as possible, and inform the user how much confidence he or she should have in the thematic information. Remote sensing–derived thematic maps should normally be subjected to a thorough accuracy assessment before being used in subsequent scientific investigations and policy decision-making (Borengasser et al., 2008; Congalton and Green, 2009). This also includes assessing the accuracy of remote sensing–derived change detection maps (Foody, 2010; Burnicki, 2011).



Overview

This chapter first introduces the steps generally taken to assess the accuracy of remote sensing–derived thematic map information (Figure 13-1). It is not concerned with the spatial (*x*, *y*) accuracy of remote sensing–derived information, which was discussed in Chapter 7. Sources of error in remote sensing–derived products are described. Error matrices (i.e., contingency tables) are introduced. The difference between training and test samples is discussed. Sample size and design are reviewed. Important information about collecting ground reference information at locations using a response design are presented. Various methods of analyzing the error matrix are discussed including qualitative evaluation and quantitative assessment. Fuzzy accuracy assessment is introduced. A method to quantify the accuracy of binary change detection maps is presented. Geostatistical analysis to assess the accuracy of remote sensing–derived information is reviewed. Finally, image metadata and lineage information that should be provided for remote sensing–derived products is discussed.

Steps to Perform Accuracy Assessment

The steps generally taken to assess the accuracy of thematic information derived from remotely sensed data are summarized in Figure 13-1. First, it is important to clearly state the nature of the thematic accuracy assessment problem, including:

- what the accuracy assessment process is expected to accomplish,
- the classes of interest (discrete or continuous), and
- the sampling design and sampling frame (consisting of area and list frames).

The objectives of the accuracy assessment should be clearly identified. Sometimes a simple qualitative examination may be appropriate if the remote sensing–derived data are to be used as general undocumented

information. Conversely, if the remote sensing–derived thematic information will impact the lives of human beings, flora, fauna, etc., then it may be necessary to conduct a thorough probability design-based accuracy assessment.

The accuracy assessment *sampling design* is the protocol (i.e., instructions) by which the reference sample units are selected (Stehman and Czaplewski, 1998; Stehman, 2009). Sampling designs generally require a *sampling frame* that consists of the materials or devices which delimit, identify, and allow access to the elements of the target population. There are two important elements associated with the sampling frame:

- the *area frame,* and
- the *list frame.*

In a remote sensing context, the area frame represents the exact geographic dimension of the entire study area. The list frame is simply a list of all the possible *sampling units* within the area frame. The actual sample is selected solely from this list of sampling units.

The accuracy assessment process is dependent upon the selection of a *sampling unit* (Stehman and Czaplewski, 1998). When conducting remote sensing–related accuracy assessment, analysts typically select one of three types of *areal* (spatial) sampling units: individual pixels (not recommended by Congalton and Green, 2009), a block of pixels (recommended), or polygons (Stehman and Wickham, 2011). There is no consensus on the ideal type of sampling unit. If a pixel is selected as the sampling unit, the land cover associated with it may be the result of a hard (crisp) classification in which each pixel is labelled as exactly one class, or a soft (fuzzy) classification in which each pixel has a membership value for each class (Stehman, 2009).

Once the accuracy assessment problem has been stated, then there are two analysis methods that may be used to validate the accuracy (or assess the error) of a remote sensing–derived thematic map:

- qualitative confidence-building assessment, and
- statistical analysis.

A *confidence-building assessment* involves the visual examination of the thematic map associated with the overall area frame by knowledgeable individuals to identify any gross errors (sometimes referred to as blunders) such as urban areas in water bodies or unrealistic classes on mountain tops. If major errors are identified and we have no qualitative confidence in the map, then the thematic information should be discarded and the classification process iterated using more appropriate logic. If the thematic map appears to be reasonable, then the analyst may proceed to statistical confidence-building measurements (Morisette et al., 2005).

Statistical analysis may be subdivided into two types (Stehman, 2000, 2001):

- model-based inference, and
- design-based inference.

Model-based inference is not concerned with the accuracy of the thematic map. It is concerned with estimating the error of the remote sensing classification *process* (or model) that actually generated the thematic map. Model-based inference can provide the user with a quantitative assessment of each classification decision. For example, each pixel in a MODIS-derived land-cover product has a confidence value that measures how well the pixel fits the training examples presented to the classifier (Morisette et al., 2005).

Design-based inference is based on statistical principles that infer the statistical characteristics of a finite population based on the sampling frame. Some common statistical measurements include producer's error, consumer's error, overall accuracy, and Kappa coefficient of agreement (Congalton and Green, 2009; Stehman, 2009). Design-based inference is expensive yet powerful because it provides unbiased map accuracy statistics using consistent estimators.

Before we describe how to extract the unbiased statistical measurements, however, it is useful to briefly review the sources of error that may be introduced into a remote sensing–derived product (e.g., a land-use or land-cover map) (Figure 13-2). This will help us appreciate why an accuracy assessment is so important when the remote sensing–derived information is used to make decisions.

Sources of Error in Remote Sensing–Derived Thematic Maps

First, error may be introduced during the remote sensing data acquisition process (Figure 13-2). It is quite common for remote sensing system detectors, cameras, RADARs, LiDARs, etc. to become miscalibrated. This can result in inaccurate remote sensing measurements (e.g., multispectral radiance, RADAR backscatter, LiDAR laser intensity). The aircraft or spacecraft platform may roll, pitch, and/or yaw during data acquisition. Erroneous commands may be entered into the data-collection or inertial navigation system from ground control personnel. Finally, the scene being imaged may be randomly affected by unwanted haze, smog, fog, dust, high humidity, or sunglint that can

General Steps Used to Assess the Accuracy of Thematic Information Derived from Remotely Sensed Data

State the nature of the thematic accuracy assessment problem.
* State what the accuracy assessment is expected to accomplish.
* Identify classes of interest (discrete or continuous).
* Specify the sampling frame within the sampling design:
  - area frame (the geographic region of interest)
  - list frame (consisting of points or areal sampling units)

Select method(s) of thematic accuracy assessment.
* Confidence-building assessment:
  - Qualitative
* Statistical measurement:
  - Model-based inference (concerned with image processing methods)
  - Design-based statistical inference of thematic information

Compute total number of observations required in the sample.
* Observations per class.

Select sampling design (scheme).
* Random.
* Systematic.
* Stratified random.
* Stratified systematic unaligned sample.
* Cluster sampling.

Obtain ground reference data at observation locations using a response design.
* Evaluation protocol.
* Labeling protocol.

Error matrix creation and analysis.
* Creation:
  - Ground reference test information (columns)
  - Remote sensing classification (rows)
* Univariate statistical analysis:
  - Producer's accuracy
  - User's accuracy
  - Overall accuracy
* Multivariate statistical analysis:
  - Kappa coefficient of agreement; conditional Kappa
  - Fuzzy

Accept or reject previously stated hypothesis.

Distribute results if accuracy is acceptable.
* Accuracy assessment report.
* Digital products.
* Analog (hard-copy) products.
* Image and map lineage report.

FIGURE 13–1 The general steps used to assess the accuracy of remote sensing–derived thematic information.

dramatically affect the quality and accuracy of the information that is extracted.

Considerable effort is spent preprocessing the remotely sensed data to remove both the geometric and radiometric error inherent in the original imagery (see Chapters 6, 7, and 12). Unfortunately, even after preprocessing, there is always residual geometric and radiometric error. For example, residual geometric error may cause pixels still to be in incorrect geographic locations. Similarly, even the best atmospheric correction will not result in a perfect relationship between percent reflectance measured on the ground and percent reflectance measured by an optical remote sensing system for the same geographic area. The residual geometric and radiometric error present in the imagery influences subsequent digital image processing (Lunetta et al., 1991; Lunetta and Lyon, 2005; Jensen et al., 2009).

Sometimes qualitative or quantitative information extraction techniques are based on flawed logic. For example, a hard land-cover classification scheme may not contain classes that are mutually exclusive, exhaustive, and hierarchical. Training sites may be incorrectly labeled during the training phase of a supervised classification. Clusters may be incorrectly labeled when performing an unsupervised classification (Chapter 9).

Sources of Error in Remote Sensing–Derived Information

Data Acquisition
- Sensor system
- Platform movement
- Ground control
- Scene considerations

Preprocessing
- Geometric correction
- Radiometric correction
- Data conversion

Information Extraction
- Classification system
- Qualitative analysis
- Quantitative analysis
- Data generalization

Data Conversion
- Raster-to-vector
- Vector-to-raster

Error Assessment
- Sampling design
- Number of samples per class
- Sample locational accuracy
- Spatial autocorrelation
- Error matrix
- Discrete multivariate statistics
- Reporting standards

Final Product Contains
- Spatial error
- Thematic error

Decisionmaking

Implement Decision

Error

FIGURE 13–2 Remote sensing–derived products such as land-use and land-cover maps contain error. The error accumulates as the remote sensing data are collected and various types of processing take place. An error assessment is necessary to identify the type and amount of error in a remote sensing–derived product (based on Figure 1 in Lunetta, R. S., Congalton, R. G., Fenstermaker, L. K., Jensen, J. R., McGwire, K. C., and L. R. Tinney, 1991, "Remote Sensing and Geographic Information System Data Integration: Error Sources and Research Issues," *Photogrammetric Engineering & Remote Sensing*, 57(6):677–687).

A human interpreter performing visual image interpretation may label an image-derived polygon incorrectly.

Thematic information derived from the remotely sensed data may be in raster (grid) or polygon cartographic data structures. Sometimes it is desirable to convert the data from one data structure to another, e.g., raster-to-vector or vector-to-raster. The conversion process may introduce error.

All the error discussed so far accumulates in the remote sensing–derived information (e.g., a land-use or land-cover map). Therefore, it is essential to conduct an error assessment (or accuracy assessment) to place confidence limits on the remote sensing–derived information. Unfortunately, the person conducting the error evaluation could make critical errors regarding the ground reference data sampling design, the number of samples obtained per class, the geometric accuracy of the sample data, and data autocorrelation. In addition, the person(s) collecting ground reference information in the field often make mistakes (e.g., mislabeling). Therefore, it is imperative to use the term *ground reference information* rather than *ground-truth* because almost all ground reference information will contain some error (Stehman and Wickham, 2011).

Hopefully, a robust multivariate statistical analysis is used so that confidence limits can be specified for the remote sensing–derived thematic information (e.g., hectares of agriculture, forestry).

The final remote sensing–derived product is created when the thematic information is judged to be sufficiently accurate. Unfortunately, the analyst can still introduce error by not following standard cartographic procedures for the creation of scale bars and map legends. Metadata may also contain error.

Finally, the people who actually make the decisions often do not understand or appreciate the amount of error that has accumulated in the remote sensing–derived thematic product (Meyer and Werth, 1990). They should be made aware of the amount of error in the product and be cautioned not to overstate the level of accuracy.

TABLE 13–1 Sample-based error matrix where there are k classes in the remote sensing–derived thematic map and N ground reference test samples that are used to assess its accuracy.

		Ground Reference Test Information Class 1 to k (j columns)					
		1	**2**	**3**		***k***	**Row total x_{i+}**
Map Class 1 to k (i rows)	**1**	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	....	$x_{1,k}$	x_{1+}
	2	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	. . .	$x_{2,k}$	x_{2+}
	3	$x_{3,1}$	$x_{3,2}$	$x_{3,3}$	. . . .	$x_{3,k}$	x_{3+}
	.	.	.	.	. . . .	.	.
	k	$x_{k,1}$	$x_{k,2}$	$x_{k,3}$	. . . .	$x_{k,k}$	x_{k+}
	Column total x_{+j}	x_{+1}	x_{+2}	x_{+3}	. . . .	x_{+k}	N

where:

- Cell entry, x_{ij}, is the proportion of area mapped as class i and labeled class j in the reference data.
- The ***row marginal***, x_{i+}, is the sum of all x_{ij} values in row i and represents the proportion of area classified as class i.
- The ***column marginal***, x_{+j}, is the sum of all x_{ij} values in column j and represents the proportion of area that is truly class j.
- The ***diagonal***, x_{ii}, summarizes correctly classified pixels.
- All off diagonal cells represent misclassified pixels.

Classification accuracy (or error) assessment was an afterthought rather than an integral part of many remote sensing studies in the 1970s and 1980s. In fact, many studies still simply report a single number (e.g., 78%) to express classification accuracy. Such *non-site-specific* accuracy assessments completely ignore locational accuracy. In other words, only the total amount of a category is considered without regard for its location. A *non-site-specific* accuracy assessment generally yields high accuracy but misleading results when all the errors balance out in a region.



The Error Matrix

To determine the accuracy of a remote sensing–derived thematic map it is necessary to systematically compare two sources of information:

1. pixels or polygons in the *remote sensing–derived classification map,* and
2. *ground reference test information* obtained at the same x, y location (which may in fact contain error).

The relationship between these two sets of information is commonly summarized in an **error matrix** (sometimes referred to as a contingency table or confusion matrix). The error matrix provides the basis on which to both describe classification accuracy and characterize errors, which may help refine the classification or the estimates derived from it (Congalton and Green, 2009; Stehman, 2009).

Stehman and Foody (2009) summarize the characteristics of a *population*-based error matrix where the analyst has reference information about the entire population within the study area (e.g., a wall-to-wall land cover reference map that can be compared to the wall-to-wall remote sensing–derived thematic map). This is a relatively rare occurrence. Most of the time it is necessary to collect samples within the study area at specific x, y locations that can be compared with those same x, y locations in the remote sensing–derived thematic map. The characteristics of a sample-based error matrix are shown in Table 13-1.

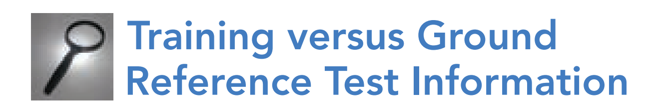
The error matrix is used to assess the remote sensing classification accuracy of the k classes in the remote sensing–derived thematic map. The central part of the error matrix is a square array of numbers $k \times k$ in size (e.g., 3×3). The columns of the matrix represent the ground reference test information and the rows correspond to the classification generated from analysis of the remotely sensed data. The intersection of the rows and columns summarizes the number of sample units (i.e., pixels, clusters of pixels, or polygons) assigned to

a particular category (class) relative to the actual category as verified in the field. The total number of samples examined is *N*.

The diagonal of the matrix (e.g., $x_{1,1}$, $x_{2,2}$) summarizes those pixels or polygons that were assigned to the correct class. Every error in the remote sensing classification relative to the ground reference information is summarized in the off-diagonal cells of the matrix (e.g., $x_{1,2}$, $x_{2,1}$, $x_{2,3}$). Each error is both an omission from the correct category and a commission to a wrong category. The column and row totals around the margin of the matrix (referred to as *marginals*) are used to compute errors of inclusion (commission errors) and errors of exclusion (omission errors). The outer row and column totals are used to compute producer's and user's accuracy (to be discussed). Some recommend that the error matrix contain proportions rather than individual counts (Stehman and Czaplewski, 1998).

But how do we a) obtain unbiased *ground reference test information* to compare with the remote sensing classification map and fill the error matrix with values, and b) perform the error (or accuracy) assessment? Basically, we must:

- be aware of the problems associated with using training versus ground reference test information;
- determine the total number of samples to be collected for each thematic category;
- design an appropriate sampling scheme (using traditional and/or geostatistical techniques);
- obtain ground reference information at sample locations using a response design; and
- apply appropriate descriptive and multivariate statistics (normal and/or fuzzy) to assess the accuracy of the remote sensing–derived information.



Training versus Ground Reference Test Information

Some image analysts perform an accuracy assessment based only on the *training pixels* (or training polygons if the study is based on human visual interpretation or object-based image analysis—OBIA) used to train classification algorithm. Unfortunately, the locations of these training sites are usually not random. They are biased by the analyst's *a priori* knowledge of where certain land-cover types exist in the scene. Because of this bias, an accuracy assessment based on how well the training class data were classified usually results in higher classification accuracies than an accuracy assessment based on unbiased ground reference test information (e.g., Muchoney and Strahler, 2002).

The ideal situation is to locate *ground reference test pixels* (or polygons) in the study area. These sites are *not* used to train the classification algorithm and therefore represent unbiased reference information. It is possible to collect some ground reference test information prior to the classification, perhaps at the same time as the training data. But the majority of test reference information is often collected after the classification has been performed using a random sample to collect the appropriate number of unbiased observations per category (to be discussed).

Landscapes often change rapidly. Therefore, it is best to collect both the training and ground reference test information as close to the date of remote sensing data acquisition as possible. This is especially important where the land use or land cover changes rapidly within a single season (Foody, 2010). For example, agriculture and illicit drug crops often have relatively short growing seasons or are harvested rapidly and then repeatedly during a season. If a scientist waits too long to obtain the ground reference test information, it may not be possible to tell what was actually in the field the day of the remote sensing data acquisition. This is a serious condition because it becomes very difficult to populate an error matrix with accurate information.

Ideally, the ground reference test data are obtained by visiting the site on the ground and making very careful observations that can then be compared with the remote sensing–derived information for the exact location (McCoy, 2005). Unfortunately, sometimes it is difficult to actually visit all the sites identified in the random sample. Some sites selected in the random sample may be completely inaccessible due to extremely rugged terrain. Others may be inaccessible because private land owners, government agencies, or even criminals (e.g., illegal drug cultivators) deny access.

When this occurs, scientists sometimes obtain higher-spatial-resolution remotely sensed data, interpret it, and use it as a surrogate for ground reference test information (e.g., Morisette et al., 2005). In such cases, the general rule of thumb is that the imagery used to obtain the ground reference test information should be substantially higher in spatial and/or spectral resolution than the imagery used to derive the original thematic information. For example, many studies have used high spatial resolution aerial photography (e.g., nominal spatial resolution <1 × 1 m) to extract ground reference test information that is then compared with thematic data produced using Landsat Thematic Mapper imagery (30 × 30 m). Admittedly, this is not the best solution. However, sometimes it is the only alternative if one wants to populate an error matrix and perform an accuracy (error) assessment.

At some point we must compare the results of the ground reference test information with the remote sensing–derived information at the appropriate location in the thematic map. When we locate a ground reference test unit in the remote sensing–derived thematic map at pixel *i, j* we do not usually just compare the pixel at that location with the ground reference test data. Rather, we often look at the 8 pixels surrounding pixel *i, j* and perhaps label the pixel according to the class that has the highest frequency of occurrence (e.g., if there are 3 corn and 5 soybean pixels in the 3 × 3 window surrounding the pixel at location *i, j*, then we would label the pixel soybean for error matrix evaluation purposes). This procedure is common practice and minimizes the effects of geometric misregistration in the remote sensing–derived product (Jensen et al., 1993).



Sample Size

The number of ground reference test samples to be used to determine the accuracy of individual categories in a remote sensing classification map is an important consideration (Foody, 2009). Usually, a design-based inferential framework is adopted in accuracy assessment programs (Stehman, 2000; 2009). With such a framework, an appropriately sized sample may be estimated using conventional statistics (Foody, 2002).

Some analysts use an equation based on the *binomial distribution* or the normal approximation to the binomial distribution to compute the required sample size. Others suggest that a *multinomial distribution* be used to determine the sample size because we are usually investigating the accuracy of multiple classes of information on a land-cover map (Congalton and Green, 2009). Alternatively, a model-based inferential framework may be adopted based on geostatistical analysis (which takes into account autocorrelation) to design an efficient sample (Stehman, 2000; Foody, 2002).

Sample Size Based on Binomial Probability Theory

Fitzpatrick-Lins (1981) suggested that the sample size N to be used to assess the accuracy of a land-use classification map be determined from the formula for the binomial probability theory:

$$N = \frac{Z^2(p)(q)}{E^2} \tag{13.1}$$

where p is the expected percent accuracy of the entire map, $q = 100 - p$, E is the allowable error, and $Z = 2$ from the standard normal deviate of 1.96 for the 95% two-sided confidence level. For a sample for which the expected accuracy is 85% at an allowable error of 5%, the number of points necessary for reliable results is (McCoy, 2005):

$$N = \frac{2^2(85)(15)}{5^2} = \text{a minimum of 204 points}.$$

With expected map accuracies of 85% and an acceptable error of 10%, the sample size for a map would be 51:

$$N = \frac{2^2(85)(15)}{10^2} = 51.$$

Thus, the lower the expected percent accuracy (p), and the greater the allowable error (E), the fewer ground reference test samples that need to be collected to evaluate the classification accuracy.

Early remote sensing research placed emphasis on achieving remote sensing-derived thematic map products that were approximately 85% (Anderson et al., 1976). Foody (2008) points out that the widely used target accuracy of 85% can be pessimistically biased and too restrictive when compared to other fields of endeavor. Pontius and Millones (2011) suggest that it is not necessary to be concerned with a universal standard for accuracy because it is not related to any specific research question or study area.

Sample Size Based on Multinomial Distribution

Analysts usually make thematic maps that contain multiple classes (e.g., vegetation, bare soil, water, urban), not just binary information (e.g., urban versus everything else). Therefore, some scientists prefer to use equations based on a *multinomial distribution* to determine the sample size necessary to assess classification accuracy. Sample size (N) derived from a multinomial distribution is based on the equation (Congalton and Green, 2009):

$$N = \frac{B\Pi_i(1 - \Pi_i)}{b_i^2} \tag{13.2}$$

where Π_i is the proportion of a population in the ith class out of k classes that has the proportion closest to 50%, b_i is the desired precision (e.g., 5%) for this class, B is the upper $(\alpha/k) \times 100$th percentile of the chi square (χ^2) distribution with 1 degree of freedom, and k is the number of classes.

For example, suppose a land-cover map contains eight classes ($k = 8$) and we know that class Π_i occupies approximately 30% of the map area and that this proportion is closest to 50%. Furthermore, we desire a level of confidence of 95% and a precision (b_i) of 5%. B is determined from the χ^2 table with 1 degree of freedom and $1-\alpha/k$ as $\chi^2_{(1,0.99375)} = 7.568$ (Congalton and Green, 2009):

$$1-\frac{\alpha}{k} = 1-\frac{0.05}{8} = 0.99375$$

$$N = \frac{7.568(0.30)(1-0.30)}{0.05^2}$$

$$N = \frac{1.58928}{0.0025} = 636 \text{ samples}$$

Therefore, in this example 636 samples should be randomly selected to adequately fill the error matrix. Approximately 80 samples per class are required (e.g., 80 × 80 = 640).

If we have no idea about the proportion of any of the classes in the land-cover map, then we can use the *worst-case* multinomial distribution algorithm where we assume that one class occupies 50% of the study area (Congalton and Green, 2009):

$$N = \frac{B}{4b^2}. \qquad (13.3)$$

Holding the precision constant at 5% for all k classes yields:

$$N = \frac{7.568}{4(0.05^2)} = 757 \text{ samples}.$$

Thus, 757 random samples would have to be obtained because we did not have prior knowledge about the true proportion of any of the k classes in this worst-case scenario.

Sometimes a confidence interval of 95% is unrealistic. For example, if we relax the confidence interval from 95% to 85%, as is standard for many land-use and land-cover mapping products as previously discussed, the new $\chi^2_{(1,0.98125)}$ value for $B = 5.695$. If we maintain the same desired precision (i.e., 5%), then the total number of samples required drops to 478:

$$1-\frac{\alpha}{k} = 1-\frac{0.15}{8} = 0.98125$$

$$N = \frac{5.695(0.30)(1-0.30)}{0.05^2}$$

$$N = \frac{1.19595}{0.0025} = 478 \text{ samples}.$$

Therefore, approximately 60 samples are required per class (e.g., 8 × 60 = 480). The worst-case scenario would require a total of 570 samples (Congalton and Green, 2009):

$$N = \frac{5.695}{4(0.05^2)} = 570 \text{ samples}$$

or approximately 71 samples per class (8 × 71 = 568).

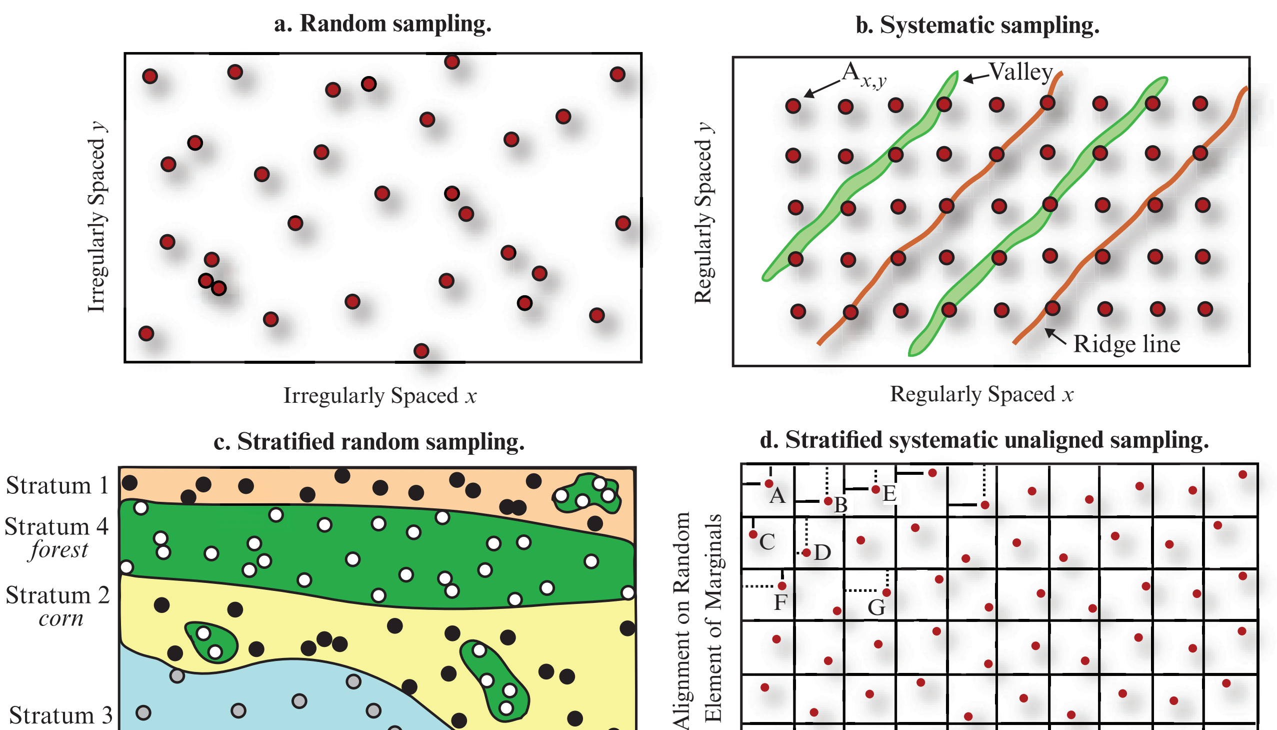
Unfortunately, it is not always possible to obtain such large numbers of random samples. A balance between what is statistically sound and what is practicably attainable must be found. Congalton (1991) and Congalton and Green (2009) suggest that a good rule of thumb is to collect a minimum of 50 samples for each land-cover class in the error matrix. Foody (2009) says this approach has a number of attractions, not least its simplicity and in the provision of a sample that allows accuracy to be assessed on a per-class basis. If the area is especially large (e.g., >1 million ha) or the classification has a large number of land-use categories (e.g., >10 classes), the minimum number of samples should be increased to 75 or 100 samples per class (McCoy, 2005). The number of samples can also be adjusted based on the relative importance of that category within the objectives of the project or by the inherent variability within each category. It may be useful to take fewer samples in categories that show little variability, such as water and forest, and increase the sampling in the categories that are more variable, such as agricultural areas. The goal is to obtain an unbiased, representative sample that can be used to populate an error matrix.



Sampling Design (Scheme)

Once the total sample size (N) and the number of samples required per class (strata) are determined, it is necessary to determine the geographic location (x, y) of these N samples in the real world so that we can visit them and obtain ground reference test information. *The sample locations must be determined randomly without bias.* Any bias will cause the statistical analysis of the error matrix to over- or underestimate the true accuracy of the thematic map. Therefore, it is necessary to select an appropriate random *sampling design* (*scheme*). Each sampling scheme assumes a different sampling model and consequently different variance equations.

There are numerous sampling designs that can be used to collect ground reference test data for assessing the accuracy of a remote sensing–derived thematic map



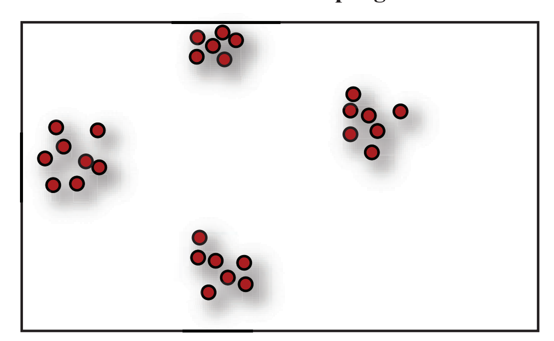


FIGURE 13–3 Several geographic sampling methods may be used to infer the statistical characteristics of a population: a) Simple random sampling. b) Systematic sampling that is initiated with random x, y coordinates selected for point A. A biased sample would likely be extracted if the systematic sample were drawn from terrain with ridge and valley topography orientation such as that shown. c) Stratified random sampling. This particular study area has been subdivided into four strata. d) Stratified systematic unaligned sampling. e) Cluster sampling.

(McCoy, 2005; Congalton and Green, 2009; Stehman, 2009). Some of the most commonly used include:

- random sampling,
- systematic sampling,
- stratified random sampling,
- stratified systematic unaligned sampling, and
- cluster sampling.

Simple Random Sampling

The key to accurately projecting the characteristics of a sample onto a population is to draw the sample with care to ensure that it is truly representative of its population (Jensen and Shumway, 2010). Simple *random sampling* without replacement generally provides adequate estimates of the population parameters, provided the sample size is sufficient (Figure 13-3a) (Congalton, 1988). Normally a random number generator is used to identify random x, y coordinates within the study area. Samples located randomly throughout a hypothetical study area are shown in Figure 13-3a. The method ensures that the members of a population are drawn independently with equal probabilities of being selected and that the selection of one sample does not influence the selection of any other sample. An advantage of working with random samples is that an important assumption underlying many statistical techniques is satisfied.

Unfortunately, random sampling may undersample small but possibly very important classes (e.g., the small acreage of illicit crops in a large agricultural region) unless the sample size is significantly large (van Oort, 2007). The random samples may also be located

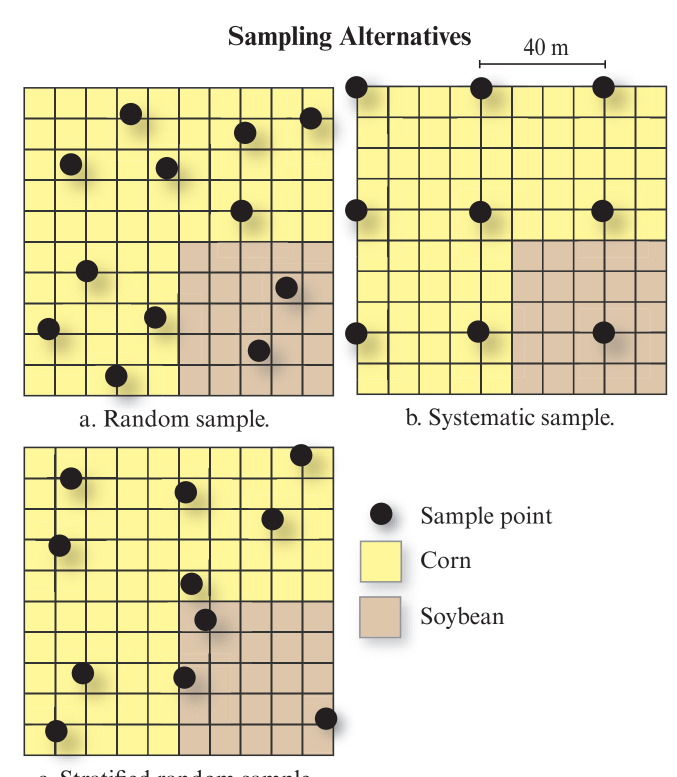


FIGURE 13–4 Hypothetical remote sensing–derived thematic map consisting of 100 pixels with 75% classified as corn and 25% classified as soybean. a) A random distribution of 12 points allocated throughout the study area. Unfortunately, only two points fall within the soybean field which should ideally be allocated three samples. Soybeans are under-sampled. b) Systematic sampling of the study area every 40 m in the *x*- and *y*-directions. Unfortunately, only one sample is allocated within the soybean field. It is undersampled. c) The random allocation of 12 samples using stratified-random sampling. As expected, 75% of the samples (9) are allocated to corn and 25% of the samples (3) are allocated to soybeans. Both received the requisite number of samples relative to the proportion of area they occupy in the classification map.

in inhospitable or access-denied locations. Finally, some clustering of the random sample points often results and, as a consequence, important spatial properties of the population may be overlooked.

A practical example of random sampling is shown in Figure 13-4a. A remote sensing–derived thematic map consisting of 100 pixels with 75% classified as corn and 25% classified as soybean is shown. Due to cost constraints, only 12 samples could be visited in the field for this particular study area. The classification of the remote sensing–derived thematic map at each sample location would be compared with the land cover found in the field at the same location. Ten of the 12 randomly located samples fall on corn (83%). Two samples fall on soybean. It is clear that soybeans are undersampled (17%) relative to the proportion of area that they occupy (25%) in the remote sensing–derived thematic map. Interestingly, collecting more random samples (e.g., 100) does not guarantee that the soybean area will ever receive any more samples.

Systematic Sampling

To avoid the shortcomings of random sampling regarding the coverage of an area, a *systematic sampling* plan may be used (Jensen and Shumway, 2010). This procedure requires that it be possible to work through the study area in some consistent and orderly manner. The sampling is normally initiated by the random selection of the starting point *x, y* coordinates (*A*) as shown in Figure 13-3b. Unfortunately, if there are periodicities in the data being collected (e.g., the ridge and valley topography of the Appalachian Mountains), the regularly spaced points could hit the same point on a cycle (e.g., a valley) time and again, yielding biased sample estimates of the population. Systematic sampling should be used with caution because it may overestimate the population parameters (Congalton, 1988). A disadvantage of both simple random sampling and systematic sampling is that the sample sizes for rare classes will be small unless the overall sample size is very large. Fortunately, both simple random sampling and systematic sampling are equal probability sampling designs, so the proportion each class represents in the sample will be approximately equal to the proportion the class represents in the region mapped (Stehman and Foody, 2009; Jensen et al., 2012).

In the hypothetical agriculture example, a systematic sample every 40 m in the *x*- and *y*-directions resulted in the allocation of only nine samples in the study area (Figure 13-4b). Eight of the nine samples were located on corn (89%). Soybeans received only one sample (11%). Decreasing the spacing of the systematic sample (e.g., to 20 m in *x*- and *y*-directions) would sample the soybean area more effectively.

Stratified Random Sampling

Analysts sometimes conduct *stratified random sampling* in which a minimum number of samples are allocated randomly to each stratum (i.e., land-cover category) after the thematic map has been prepared (Jensen and Shumway, 2010). For example, a stratified random sample based on 10 geographic regions was used to assess the accuracy of the U.S. National Land Cover Dataset (NLCD) and impervious surface (Wickham et al., 2010, 2013).

Stratified random sampling involves two steps. First, the study area is subdivided into land-cover strata based on what is found in the remote sensing–derived

thematic map. A stratum is created by extracting only pixels (or polygons) associated with a specific class. Some analysts prefer to mask out all unwanted classes to create the stratum file. Sample locations are then randomly distributed throughout the geographic strata. For example, all pixels (or polygons in a visual interpretation project) associated with forest were placed in a single stratum (4) shown in Figure 13-3c. A random number generator was then used to allocate a sufficient number of random samples to this forest stratum. Similarly, all pixels (or polygons) associated with corn cultivation were placed in stratum 2 in Figure 13-3c. Random samples were then allocated throughout this stratum. Note that only strata 3 and 4 were completely homogeneous in Figure 13-3c. Strata 1 and 2 had stratum 4 (forest) areas within them. This is quite common. Note that the stratification process resulted in random samples being allocated to the forest islands even though they existed within other strata.

The advantage of stratified random sampling is that all strata (e.g., land-cover classes), no matter how small their proportion of the entire study area, will have samples allocated to them for error evaluation purposes. Without stratification, it is often difficult to find sufficient samples for classes that occupy a very small proportion of the study area. The drawback with stratified random sampling is that one has to wait until the thematic map has been completed in order to allocate samples to the various land-cover strata. The ground reference test information can rarely be obtained on the same day as the remote sensing data acquisition.

Stratified random sampling can also be applied to products derived from remote sensing area-frame sampling as long as the various land-cover field polygon boundaries associated with the individual frames are kept. This allows ground reference test samples to be allocated within each stratum, as discussed (Wickham et al., 2010, 2013; Stehman et al., 2012).

In the hypothetical agriculture example, stratified random sampling allocated 75% of the samples (9) within the corn region and 25% of them (3) within the soybean region as expected (Figure 13-4b). Both were allocated the requisite number of samples relative to the proportion of area they occupy in the study area.

Stratified Systematic Unaligned Sampling

S*tratified systematic unaligned sampling* combines randomness and stratification with a systematic interval. It introduces more randomness than just beginning with a random *x, y* coordinate for the first sample in each stratum. Figure 13-3d demonstrates how a stratified systematic unaligned sample is created. First, point *A* is selected at random. The *x*-coordinate of *A* is then used with a new random *y*-coordinate to locate *B*, a second random *y*-coordinate to locate *E*, and so on across the top *row* of the stratum under investigation. By a similar process the *y*-coordinate of *A* is used in combination with random *x*-coordinates to locate point C and all successive points in the first *column* of the stratum under investigation. Then, the random *x*-coordinate of *C* and *y*-coordinate of *B* are used to locate *D*, of *E* and *F* to locate *G,* and so on until all strata have sample elements.

Cluster Sampling

Sometimes it is difficult to go into the field to randomly selected locations and obtain the required information. Therefore, some have suggested that several samples can be collected at a single random location. Unfortunately, each pixel in the cluster is not independent of the others. Therefore, Congalton (1988) warns that clusters no larger than 10 pixels and certainly not larger than 25 pixels should be used because of the lack of information added by each pixel beyond these cluster sizes.

Some combination of random and stratified sampling provides the best balance of statistical validity and practical application. Such a system may employ random sampling to collect some assessment data early in the project, and random sampling within strata should be used after the classification has been completed to ensure that enough samples were collected for each category and to minimize periodicity (spatial autocorrelation) in the data (Congalton, 1988). Ideally, the *x, y* locations of the reference test sites are determined using global positioning system (GPS) instruments.

Charleston, SC, Case Study: To demonstrate several of the concepts, consider the 407 reference pixels used to assess the accuracy of the land-cover map of Charleston, SC, derived from the Landsat Thematic Mapper imagery shown in Figure 9-18a. The accuracy was assessed based on stratified random sampling after classification. The methods involved making five files, each containing only the pixels of a specific land-cover class (i.e., the classified land-cover pixels in each file were recoded to a value of 1 and the unclassified background to a value of 0). A random number generator was then used to identify random *x, y* coordinates within each stratified file until a sufficient number of points was collected (i.e., at least 50 for each class). The result was a stratified random sample of the five classes. All locations were then visited in the field or evaluated using large-scale orthophotography acquired during the same month as the Landsat TM data collection. Cross-tabulating the information resulted in the creation of the error matrix shown in Table 13-2.

TABLE 13–2 Stratified random sample-based error matrix associated with the land cover classification map derived from Landsat Thematic Mapper data of Charleston, SC, shown in Figure 9-18a. The map has five classes and 407 ground reference samples were collected.

		Ground Reference Test Information Class 1 to k (j columns)					
		Residential	Commercial	Wetland	Forest	Water	Row total x_{i+}
Map Class 1 to k (i rows)	Residential	**70**	5	0	13	0	88
	Commercial	3	**55**	0	0	0	58
	Wetland	0	0	**99**	0	0	99
	Forest	0	0	4	**37**	0	41
	Water	0	0	0	0	**121**	121
	Column total x_{+j}	73	60	103	50	121	**407**

Overall Accuracy = 382/407 = 93.86%

Producer's Accuracy (omission error)

Residential = 70/73 =	96%	4% omission error
Commercial = 55/60 =	92%	8% omission error
Wetland = 99/103 =	96%	4% omission error
Forest = 37/50 =	74%	26% omission error
Water = 20/22 =	100%	0% omission error

User's Accuracy (commission error)

Residential = 70/88 =	80%	20% commission error
Commercial = 55/58 =	95%	5% commission error
Wetland = 99/99 =	100%	0% commission error
Forest = 37/41 =	90%	10% commission error
Water = 121/121 =	100%	0% commission error

Computation of K_{hat} Coefficient of Agreement:

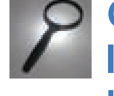
$$\hat{K} = \frac{N\sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k}(x_{i+} \times x_{+j})}{N^2 - \sum_{i=1}^{k}(x_{i+} \times x_{+j})}$$

where N = 407

$$\sum_{i=1}^{k} x_{ii} = (70 + 55 + 99 + 37 + 121) = 382$$

$$\sum_{i=1}^{k}(x_{i+} \times x_{+j}) = (88 \times 73) + (58 \times 60) + (99 \times 103) + (41 \times 50) + (121 \times 121) = 36{,}792$$

$$\text{therefore } \hat{K} = \frac{407(382) - 36792}{407^2 - 36792} = \frac{155474 - 36792}{165649 - 36792} = \frac{118682}{128857} = 92.1\%$$



Obtaining Ground Reference Information at Locations Using a Response Design

The actual collection of ground reference samples for use in the accuracy assessment is technically referred to as the *response design* (Stehman and Czaplewski, 1998; Wickham et al., 2010). In effect, it is the data collection conducted in response to the sampling design. The response design can be subdivided into an evaluation protocol and a labeling protocol. The *evaluation protocol* refers to selecting the support region on the ground (typically associated with a pixel or polygon) where the

ground information will be collected. Atkinson and Curran (1995) define this as the spatial support region consisting of "the size, geometry and orientation of the space on which an observation is defined." Once the location and dimensions of the sampling unit are defined, the *labeling protocol* is initiated and the sampling unit is assigned a hard or fuzzy ground reference label. This ground reference label (e.g., forest) is paired with the remote sensing–derived label (e.g., forest) for assignment in the error matrix.



Evaluation of Error Matrices

After the ground reference test information has been collected from the randomly located samples, the test information is compared pixel by pixel (or polygon by polygon) with the information in the remote sensing–derived classification map. Agreement and disagreement are summarized in the cells of the error matrix as demonstrated in Table 13-2. Information in the error matrix may be evaluated using simple descriptive statistics or multivariate statistical techniques.

Descriptive Evaluation of Error Matrices

There is no agreement on a standard approach to thematic map accuracy assessment based on the analysis of error matrices. Each accuracy measure may be more relevant than others for a particular objective because different accuracy measures incorporate different information about the error matrix. The use of different accuracy measures may result in different, possibly even conflicting, interpretations and conclusions (Foody, 2002; Liu et al., 2007).

Liu et al. (2007) summarized 34 measures that were applied to 595 error matrices found in the literature to assess classification accuracy. Their multidimensional scaling analysis provided detailed information about the suitability of the various measures. Only three heavily used measures are discussed here (e.g., Wickham et al., 2010, 2013).

Nomenclature associated with a sample-based error matrix was presented in Table 13-1. The *overall accuracy* of the classification map is determined by dividing the total number of correctly classified pixels (i.e., the sum of the diagonal in the error matrix) by the total number of pixels in the error matrix (N) (Story and Congalton, 1986; Congalton and Green, 2009):

$$\text{Overall accuracy} = \frac{\sum_{i=1}^{k} x_{ii}}{N}. \qquad (13.4)$$

Liu et al. (2007) found that this should be the primary measure for assessing the accuracy of an error matrix.

Computing the accuracy of individual classes, however, is more complex because the analyst has the choice of dividing the number of correct pixels in the class by the total number of pixels in the corresponding row or column. The total number of correct pixels in a class divided by the total number of pixels of that class as determined from the ground reference data (i.e., the column total) is called the *producer's accuracy* (Story and Congalton, 1986; Congalton and Green, 2009):

$$\text{Producer's accuracy for class } j = \frac{x_{jj}}{x_{+j}}. \qquad (13.5)$$

This statistic indicates the probability of a reference pixel being correctly classified and is a measure of omission error.

If the total number of correct pixels in a class is divided by the total number of pixels that were actually classified in that category, the result is a measure of commission error. This measure, called the *user's accuracy* or reliability (Story and Congalton, 1986; Congalton and Green, 2009):

$$\text{User's accuracy for class } i = \frac{x_{ii}}{x_{i+}} \qquad (13.6)$$

is the probability that a pixel classified on the map actually represents that category on the ground.

Sometimes we are producers of classification maps and sometimes we are users. Therefore, it is good practice to report all three accuracy measures: overall accuracy, producer's accuracy, and user's accuracy, because we never know how the classification will be used (Lu and Weng, 2007; Wickham et al., 2013; Pouliot et al., 2014).

For example, the remote sensing–derived error matrix in Table 13-2 has an overall classification accuracy of 93.86%. However, what if we were primarily interested in the ability to classify just residential land use using Landsat TM data of Charleston, SC? The producer's accuracy for this category was calculated by dividing the total number of correct pixels in the category (70) by the total number of residential pixels as indicated by the reference data (73), yielding 96%, which is quite good. We might conclude that because the overall accuracy of the entire classification was 93.86% and the producer's accuracy of the residential land-use class was 96%, that the procedures and Landsat TM data used are adequate for identifying residential land use in this area. Such a conclusion could be a mistake. We should not forget the user's accuracy, which is computed by dividing the total number of correct pixels in the

residential category (70) by the total number of pixels classified as residential (88), yielding 80%. In other words, although 96% of the residential pixels were correctly identified as residential, only 80% of the areas called residential are actually residential.

A careful evaluation of the error matrix also reveals that there was confusion when discriminating residential land use from commercial and forest land cover. Therefore, although the producer of this map can claim that 96% of the time a geographic area that was residential was identified as such, a user of this map will find that only 80% of the time will an area she or he visits in the field using the map will actually be residential. The user may feel that an 80% user's accuracy is unacceptable.

Discrete Multivariate Techniques Applied to the Error Matrix

Discrete multivariate techniques have been used to evaluate the accuracy of remote sensing–derived classification maps and error matrices since 1983 (Congalton and Mead, 1983). The techniques are appropriate because remotely sensed data are discrete rather than continuous and are also binomially or multinomially distributed rather than normally distributed. Statistical techniques based on normal distributions simply do not apply. It is important to remember, however, that there is not a single universally accepted measure of accuracy, but instead there are a variety of indices, each sensitive to different features (Foody, 2002; Lu and Weng, 2007).

It is useful to review several multivariate error evaluation techniques using the error matrix in Table 13-2. First, the raw error matrix can be *normalized* (standardized) by applying an iterative proportional fitting procedure that forces each row and column in the matrix to sum to 1 (not shown). In this way, differences in sample sizes used to generate the matrices are eliminated and individual cell values within the matrix are directly comparable. In addition, because the rows and columns are totaled (i.e., the marginals) as part of the iterative process, the resulting normalized matrix is more indicative of the off-diagonal cell values (i.e., the errors of omission and commission). In other words, all the values in the matrix are iteratively balanced by row and column, thereby incorporating information from that row and column into each cell value. This process then changes the cell values along the major diagonal of the matrix (correct classification), and therefore a normalized overall accuracy can be computed for each matrix by summing the major diagonal and dividing by the total of the entire matrix. Therefore, it may be argued that the normalized overall accuracy is a better representation of accuracy than is the overall accuracy computed from the original matrix because it contains information about the off-diagonal cell values (e.g., Congalton, 1991; Stehman and Foody, 2009).

Standardized error matrices are valuable for another reason. Consider a situation where analyst 1 uses classification algorithm A and analyst 2 uses classification algorithm B on the same study area to extract the same four classes of information. Analyst A evaluates 250 random locations to derive error matrix A and analyst B evaluates 300 random locations to derive error matrix B. After the two error matrices are standardized, it is possible to directly compare cell values of the two matrices to see which of the two algorithms was better. Therefore, the normalization process provides a convenient way to compare individual cell values between error matrices regardless of the number of samples used to derive the matrix.

Kappa Analysis

Kappa analysis is a discrete multivariate technique of use in accuracy assessment. The method was introduced to the remote sensing community in 1981 and was first published in a remote sensing journal in 1983 (Congalton, 1981; Congalton et al., 1983).

K_{hat} Coefficient of Agreement Kappa computes a statistic, $\hat{K}$, which is an estimate of Kappa. It is a measure of agreement or accuracy between the remote sensing–derived classification map and the reference data as indicated by a) the major diagonal, and b) the chance agreement, which is indicated by the row and column totals (often referred to as *marginals*) (Paine and Kiser, 2003; Congalton and Green, 2009). $\hat{K}$ is computed:

$$\hat{K} = \frac{N\sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k}(x_{i+} \times x_{+j})}{N^2 - \sum_{i=1}^{k}(x_{i+} \times x_{+j})} \qquad (13.7)$$

where k is the number of rows (e.g., land-cover classes) in the error matrix, x_{ii} is the number of observations in row i and column i, and x_{i+} and x_{+j} are the marginal totals for row i and column j, respectively, and N is the total number of samples. $\hat{K}$ values >0.80 (i.e., >80%) represent strong agreement or accuracy between the classification map and the ground reference information. $\hat{K}$ values between 0.40 and 0.80 (i.e., 40% to 80%) represent moderate agreement. $\hat{K}$ values <0.40 (i.e., <40%) represent poor agreement.

Congalton and Green (2009) caution that if a standardized error matrix is used that has a great many off-diagonal cells with zeros (i.e., the classification is very

good), then the normalized results may disagree with the overall classification accuracy and standard Kappa results.

Charleston, SC, Case Study: The computation of the $\hat{K}$ statistic for the Charleston, SC, dataset is summarized in Table 13-2. The overall classification accuracy is 93.86%, and the $\hat{K}$ statistic is 92.1%. The results are different because the two measures incorporated different information. The overall accuracy incorporated only the major diagonal and excluded the omission and commission errors. Conversely, $\hat{K}$ computation incorporated the off-diagonal elements as a product of the row and column marginals. Therefore, depending on the amount of error included in the matrix, these two measures may not agree. Congalton (1991) suggests that overall accuracy, normalized accuracy, and $\hat{K}$ be computed for each matrix to "glean as much information from the error matrix as possible." Computation of the $\hat{K}$ statistic may also be used 1) to determine whether the results presented in the error matrix are significantly better than a random result (i.e., a null hypothesis of $\hat{K} = 0$), or 2) to compare two similar matrices (consisting of identical categories) to determine if they are significantly different.

Savannah River Site, SC, Case Study: It is important to monitor the land cover of clay-capped hazardous waste sites on a repetitive basis to ensure the integrity of the clay caps (Jensen et al., 2003). This was accomplished at the Westinghouse Savannah River Site near Aiken, SC, using DAIS 3715 hyperspectral imagery obtained on July 31, 2002, and a Spectral Angle Mapper classification algorithm. The accuracy of the classification was determined by first collecting detailed *in situ* land-cover reference data at the 98 locations shown in Figure 13-5a. All test sites were inventoried to within ± 30 cm using survey-grade GPS. The land cover at each of these locations was identified in the classification map shown in Figure 13-5b. The classification accuracy results are summarized in Table 13-3. The overall classification accuracy was 89.79% with a $\hat{K}$ statistic of 85.81%.

Conditional K_{hat} Coefficient of Agreement The conditional coefficient of agreement ($\hat{K}_c$) can be used to calculate agreement between the reference and remote sensing–derived data with chance agreement eliminated for an individual class for user accuracies using the equation (Paine and Kiser, 2003; Congalton and Green, 2009):

$$\hat{K}_i = \frac{N(x_{ii}) - (x_{i+} \times x_{+j})}{N(x_{i+}) - (x_{i+} \times x_{+j})} \quad (13.8)$$

where x_{ii} is the number of observations correctly classified for a particular category (summarized in the diagonal of the matrix), x_{i+} and x_{+j} are the marginal totals for row i and column i associated with the category, and N is the total number of observations in the entire error matrix. For example, the conditional $\hat{K}_i$ coefficient of agreement for the residential land-use class of the Charleston, SC, dataset (Table 13-2) is:

$$\hat{K}_{Resid} = \frac{407(70) - (88 \times 73)}{407(88) - (88 \times 73)} = \frac{28490 - 6424}{35816 - 6424} = 0.75$$

This procedure can be applied to each land-cover class of interest.

Kappa Debate Hundreds, perhaps even thousands of remote sensing investigations have used Kappa analysis to assess the accuracy of remote sensing–derived land-cover maps. It is important to note, however, that some scientists do not believe that Kappa analysis is the most appropriate measure for assessing the accuracy of a remote sensing–derived thematic map. For example, Liu et al. (2007) evaluated 34 measures applied to error matrices and did not recommend the use of Kappa measures. Foody (2009) suggested that some of the arguments offered for the use of Kappa may be flawed and/or apply equally to other measures of accuracy. Pontius and Millones (2011) evaluated five different Kappa indices including some put forth by Pontius. They suggested that the Kappa indices are misleading and/or flawed for practical applications. They then proposed the analysis of error matrices using two summary parameters: quantity disagreement and allocation.

Conversely, Congalton (2005) and Congalton and Green (2009) believe that Kappa analysis has become a standard component of most accuracy assessments and is considered a required component of most digital image processing software programs. The remote sensing community does not always adopt every new accuracy assessment approach that is put forth. It will be interesting to see what direction the remote sensing community takes.

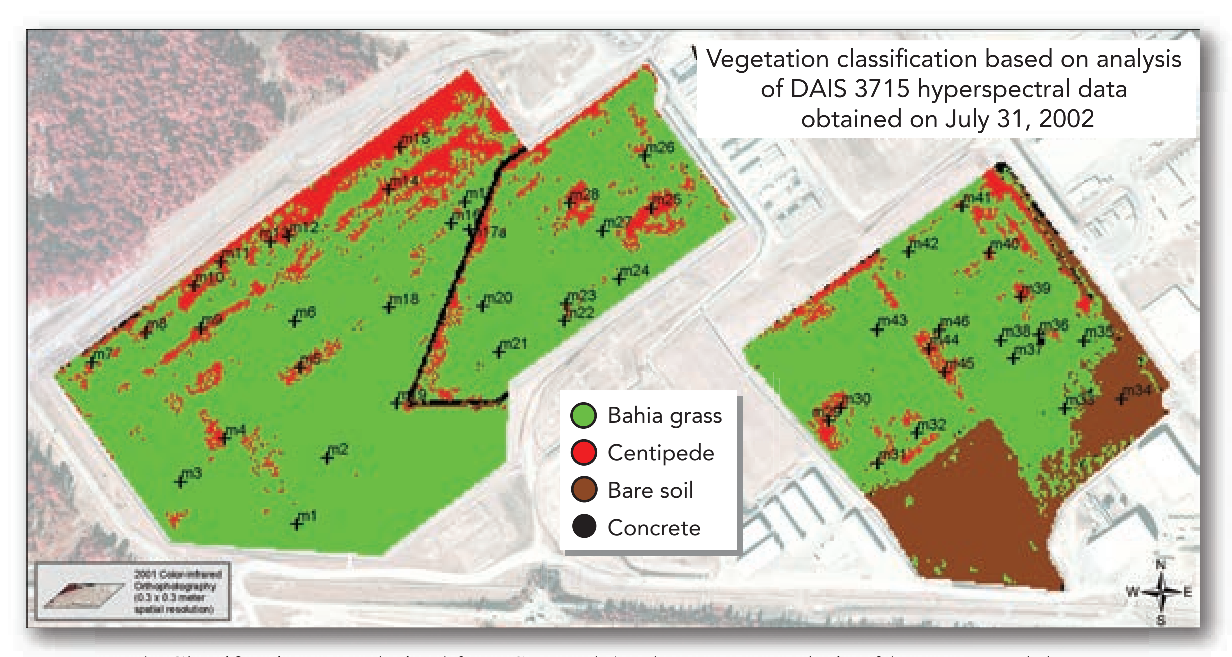
Fuzzification of the Error Matrix

A hard classification system requires that all classes be mutually exclusive and exhaustive. Furthermore, the classes must usually be hierarchical, meaning that more detailed classes logically can be combined to produce more general but useful classes. Some phenomena in the real world may in certain instances be considered pure or homogeneous (e.g., a monoculture corn field with complete canopy closure). Unfortunately, the world contains many phenomena that are not well mapped using hard classification logic. Sometimes a field or polygon contains multiple materials. These are

Accuracy Assessment In Situ Measurement Locations and Remote Sensing–Derived Classification Map of the Mixed Waste Management Facility on the Savannah River Site



a. *In situ* measurement locations.



b. Classification map derived from Spectral Angle Mapper analysis of hyperspectral data.

FIGURE 13–5 Classification accuracy assessment of a land-cover map of the Mixed Waste Management Facility on the Savannah River Site in 2002 based on the analysis of DAIS 3715 hyperspectral data and *in situ* ground reference data. a) The location of Bahia grass and Centipede grass *in situ* measurement locations. b) Remote sensing–derived classification map.

often referred to as endmembers (e.g., corn intercropped with hay, bare soil between rows of corn, shadows, etc.). In this case, per-pixel or polygon class membership may be a matter of degree or percentage rather than assignment to a hard, mutually exclusive class (Foody, 2002). For example, if a single field or polygon contains 49% corn and 51% hay, should it really be labeled a hay field or should it be called something else? In fact, many natural landscapes often grade gradually into one another at the interface (e.g., it is common to find grassland grading into scrub/shrub and then scrub/shrub transitioning into forest).

TABLE 13–3 Accuracy assessment of the classification map derived from hyperspectral data of the Mixed Waste Management Facility on the Savannah River Site (Jensen et al., 2003).

		Ground Reference Test Information Class 1 to k (j columns)				
		Bahia grass	**Centipede**	**Bare soil**	**Concrete**	**Row total x_{i+}**
Map Class 1 to k (i rows)	**Bahia grass**	**31**	2	0	0	33
	Centipede	6	**7**	0	2	15
	Bare soil	0	0	**30**	0	30
	Concrete	0	0	0	**20**	20
	Column total x_{+j}	37	9	30	22	**98**

Overall Accuracy = 88/98 = 89.79%

Producer's Accuracy (omission error)
Bahia grass = 31/37 = 84% 16% omission error
Centipede = 7/9 = 78% 22% omission error
Bare soil = 30/30 = 100% 0% omission error
Concrete = 20/22 = 91% 9% omission error

User's Accuracy (commission error)
Bahia grass = 31/33 = 94% 6% commission error
Centipede = 7/15 = 47% 53% commission error
Bare soil= 30/30 = 100% 0% commission error
Concrete = 20/20 = 100% 0% commission error

Computation of K_{hat} Coefficient of Agreement

$$\hat{K} = \frac{N \sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k} (x_{i+} \times x_{+j})}{N^2 - \sum_{i=1}^{k} (x_{i+} \times x_{+j})}$$

where $N = 98$

$$\sum_{i=1}^{k} x_{ii} = (31 + 7 + 30 + 20) = 88$$

$$\sum_{i=1}^{k} (x_{i+} \times x_{+j}) = (33 \times 37) + (15 \times 9) + (30 \times 30) + (20 \times 22) = 2696$$

therefore $\hat{K} = \frac{98(88) - 2696}{98^2 - 2696} = \frac{8624 - 2696}{9604 - 2696} = \frac{5928}{6908} = 85.81\%$

Gopal and Woodcock (1994) were among the first to suggest that fuzzy logic be used to introduce real-world fuzziness into the classification map accuracy assessment process. Instead of a right or wrong (binary) analysis, map labels were considered partially right (or wrong), generally on a five-category scale. Various statistics were then derived from the fuzzy information, including Max (M), the number of sites with an absolutely right answer (rank of 5); Right (R), the number of sites with a reasonable, good, or absolutely right answer (ranks 3, 4, and 5); and Increase (R–M), which is the difference between the Right and Max functions (Jacobs and Thomas, 2003). These were used to produce more sophisticated fuzzy logic–based accuracy assessment measures.

NIMA GeoCover Land-cover Case Study: Green and Congalton (2003) used fuzzy logic during the phase

Ground Reference Polygon: 21

Analyst: Ryan

Date: 00.00.00

Comments: Appears to be a fuzzy transition from primarily deciduous forest into some evergreen forest.

Reference material used: Landsat Thematic Mapper 30 × 30 m imagery. National Technical Means imagery. Best available large-scale topographic maps.

Back Next

Fuzzy Accuracy Assessment

Classification	Good	Acceptable	Poor
Deciduous forest	●	○	○
Evergreen forest	○	●	○
Scrub/shrub	○	○	●
Grass	○	○	●
Barren	○	○	●
Urban	○	○	●
Agriculture	○	○	●
Water	○	○	●

FIGURE 13–6 A template used for the introduction of fuzzy reference test information in an accuracy assessment of NIMA-sponsored Landsat GeoCover land-cover maps. The ground reference test polygon under consideration (21) has been assigned the deterministic class of deciduous forest. However, the analyst believes it could also be evergreen forest. The analyst is sure it is not one of the other categories. These data can be used to create an error matrix that contains deterministic as well as fuzzy accuracy assessment information (based on Green and Congalton, 2003).

when ground reference information were collected and compared it with the remote sensing–derived classified map results. Their goal was to assess the accuracy of a NIMA-sponsored global Landsat (30 × 30 m) GeoCover land-cover map (Green and Congalton, 2003). It was not practical to visit on the ground all the sites selected in the worldwide stratified random sample. Therefore, NIMA provided high spatial resolution national technical means (NTM) imagery of the selected sample sites. These data were interpreted and represented the ground reference test information.

Experienced interpreters used a version of the fuzzy accuracy assessment form shown in Figure 13-6 to enter their ground reference test information. Each ground reference test sample location in the NTM imagery was evaluated for the likelihood of being identified as each of the possible cover types. The analyst first determined the most appropriate label for the sample site, e.g. deciduous forest, which was then entered in the "classification" column in the form (Figure 13-6). After assigning the deterministic label for the sample, the remaining possible map labels were evaluated as "good," "acceptable," or "poor" candidates for the site's label. For example, a sample site might fall near the classification scheme margin between deciduous forest and evergreen forest. In this instance, the analyst might rate evergreen forest as acceptable, but scrub/shrub and the other categories as "poor." As each site was interpreted, the deterministic and fuzzy assessment reference labels were entered into the error matrix (Table 13-4) (Green and Congalton, 2003).

Nondiagonal cells in the error matrix (Table 13-4) contain two tallies that can be used to distinguish class labels that are uncertain or that fall on class margins from class labels that are probably in error (Green and Congalton, 2003; 2005). The first number represents sites where the map label matched a "good" or "acceptable" reference label in the fuzzy assessment. Therefore, even though the class label was not considered the most appropriate, it was considered acceptable given the fuzziness of the classification system and the minimal quality of some of the ground reference test data. These sites are considered a "match" for estimating fuzzy assessment accuracy. The second number in the cell represents those sites whose map class label was considered poor (i.e., an error).

The fuzzy assessment overall accuracy was estimated as the percentage of sites where the "good" or "acceptable" reference label(s) matched the map label. Individual class accuracy was estimated by summing the number of matches for that class's row or column divided by the row or column total. Class accuracy by column represented producer's class accuracy. Class accuracy by row represented user's accuracy (Green and Congalton, 2003; Congalton and Green, 2009).

TABLE 13–4 An error matrix that contains deterministic and fuzzy information used to assess the classification accuracy of NIMA-sponsored Landsat GeoCover thematic map products derived from Landsat imagery (based on Green and Congalton, 2003).

	Ground Reference Data Class 1 to *k* (*j* columns)									
	D. Forest	**E. Forest**	**Scrub/ Shrub**	**Grass**	**Barren**	**Urban**	**Ag.**	**Water**	**User's Row Total Deterministic**	**User's Row Total Fuzzy**
Deciduous Forest	**48**	24,7	0,1	0,3	0,0	0,1	0,11	0,18	48/113 (42.5%)	72/113 (63.7%)
Evergreen Forest	4,0	**17**	0,1	0,0	0,0	0,0	0,1	0,3	17/26 (65.4%)	21/26 (80.8%)
Scrub/shrub	2,0	0,1	**15**	8,1	0,0	0,0	2,2	0,0	15/31 (48.4%	27/31 (87.1%)
Grass	0,1	0,0	5,1	**14**	0,0	0,0	3,0	0,0	14/24 (58.3%)	22/24 (91.7%)
Barren	0,0	0,0	0,2	0,0	**0**	0,0	0,1	0,0	0/3 (0%)	0/3 (0%)
Urban	0,0	0,0	0,0	0,0	0,0	**20**	2,0	0,0	20/22 (90.9%)	22/22 (100%)
Agriculture	0,1	0,1	7,15	18,6	0,0	2,0	**29**	1,2	29/82 (35.4%)	57/82 (69.5%)
Water	0,0	0,0	0,0	0,0	0,0	0,0	0,0	**8**	8/8 (100%)	8/8 (100%)
Producer's Column Total Deterministic	48/56 (85.7%)	17/50 (34%)	15/47 (31.9%)	14/50 (28%)	na na	20/24 (83.3%)	29/51 (56.9%)	8/33 (24.2%)		
Producer's Column Total Fuzzy	54/56 (96.4%)	41/50 (82%)	27/47 (57.4%)	40/50 (80%)	na na	22/24 (91.7%)	36/51 (70.6%)	10/33 (33.3%)		
Overall deterministic accuracy = 151/311 = 48.6% **Overall fuzzy accuracy** = 230/311 = 74%										

Overall accuracy for the deterministic components of the matrix was 48.6% (151/311). The deterministic statistic ignored any variation in the interpretation of reference data and the inherent fuzziness at class boundaries. Including the "good" and "acceptable" ratings, overall fuzzy accuracy was 74% (230/311).

Gopal and Woodcock (1994) introduced groundbreaking methods for quantifying accuracy when the reference classification is fuzzy and the map classification is hard (crisp) (Stehman, 2009). Pontius and Cheuk (2006) extended the error matrix concept to provide descriptive accuracy measures when both the map and the reference classifications are fuzzy. Gomez et al. (2008) developed new accuracy measures that provide overall, producer's, user's or Kappa statistics when working with fuzzy or hard (crisp) error matrices.

An error matrix and the estimates of classification accuracy derived from it provide no information on the spatial distribution of the error (Foody, 2005). Comber et al. (2012) used geographically weighted regression approaches to model the spatial variations in the accuracy of both hard (crisp) Boolean and soft (fuzzy) land-cover classes.



Change Detection Map Accuracy Assessment

van Oort (2007) and Warner et al. (2009) suggest that there are three basic types of error matrices reported for change detection products:

1. Accuracy assessment of the classification of each independent date.

TABLE 13–5 Hypothetical accuracy assessment of a remote sensing–derived thematic map with just three classes: Buildings, Roads, and Other (based on Jensen et al., 2012).

		Ground Reference Information Class 1 to *k* (*j* columns)			
		Building	**Road**	**Other**	**Row total**
Map Class 1 to *k* (*i* rows)	**Building**	**49**	4	5	58
	Road	5	54	4	63
	Other	6	**6**	**59**	71
	Column total	60	64	68	**192**
	Overall accuracy = 162/192 = 84.4%				
	Producer's accuracy	Buildings = 49/60 = 82%	Roads = 54/64 = 84%	Other = 59/68 = 87%	
	User's accuracy	Buildings = 49/58 = 82%	Roads = 54/63 = 84%	Other = 59/71 = 87%	
	Kappa Coefficient of Agreement (κ) = [192(162) – 12,340] / (36864 – 12,340) = 76.5%				

2. A complete change transition error matrix consisting of all possible change and no-change classes and their potential confusion with each other.
3. Accuracy assessment of a simple binary change/no-change classification.

Assessing the Accuracy of the Individual Thematic Maps used in a Change Detection Study

This type of accuracy assessment is performed for each of the classification maps used in the change detection study. Table 13-5 demonstrates how a simple 3 × 3 error matrix is sufficient for assessing the accuracy of a single-date three-class map consisting of Building, Road, and Other land cover. Unfortunately, analysis of the accuracy of the individual date thematic maps does not provide any accuracy information about the change map produced from the intersection of the two individual date thematic maps.

Assessing the Accuracy of a "From–To" Change Detection Map

Assessing the accuracy of a change detection map consisting of "from–to" classes is no trivial matter (Khorram et al., 1999; Foody, 2002; Jensen and Im, 2007). For example, when evaluating the thematic accuracy of a change detection map with the same three classes (Building, Road, Other) derived from two dates of imagery, there is now the possibility of 81 change classes in the change map as shown in Table 13-6. Thus, if there are k classes in each of the maps used to detect change, then the total number of change classes equals $k^2 \times k^2$ (Warner et al., 2009). This significantly complicates the process of collecting a sufficient number of ground reference samples to populate the error matrix (Congalton and Green, 2009; Warner et al., 2009).

Response Design

The response design considerations (e.g., per pixel, polygon) used to assess the accuracy of a single-date thematic map hold true when analyzing the information in a change map (Jensen et al., 2012). However, the user must be particularly careful about specifying the desired acceptable level of accuracy for change detection map products. They may not be the same as those used when assessing the accuracy of a single date thematic map.

Sampling Design

The same general rule-of-thumb about the required number of samples unfortunately holds true when assessing the accuracy of a change map, i.e., approximately 50 samples per class. However, when assessing the accuracy of a simple 3-class change map, the analyst must now somehow determine the land cover on 450 random locations (9 change classes × 50 samples per class) (Congalton and Green, 2009). This is a daunting task. Also, some of the possible "from–to" change classes have a very low probability of occurring, such as a change from "Building" to "Other" unless the building was destroyed. Nevertheless, the methods for obtaining the requisite reference samples to document the accuracy of the change detection map

TABLE 13–6 A change detection error matrix for the same three map categories [Building (B), Road (R), and Other (O)] that were used in the single-date accuracy assessment (based on Macleod and Congalton, 1998; Jensen et al., 2012).

		Ground Reference Information Class 1 to *k* (*j* columns)									
		BB	RR	OO	BR	BO	RB	RO	OB	OR	Row total
Change Detection Map Class 1 to *k* (*i* rows)	BB	|									
	RR										
	OO			|							
	BR										
	BO										
	RB										
	RO										
	OB								|		
	OR								|		
	Column total										N

are the same as those used when assessing the accuracy of an individual thematic map, including: random sampling, stratified sampling, systematic sampling, and cluster sampling (Jensen et al., 2012).

Analysis

Given careful planning and sufficient resources to obtain sufficient ground reference points at two dates in time, it is possible to populate the change detection error matrix with measurements. The major diagonal of the change matrix is highlighted in black in Table 13-6 and represents when the Date 1 classification of a pixel or polygon agrees with the Date 2 classification of the same pixel or polygon (e.g., it was a "Building" on Date 1 and a "Building" on Date 2 [i.e., BB]). Similarly, if the pixel or polygon on Date 1 was "Other" and "Other" on Date 2, then a hash mark would be placed in the OO cell as shown.

The matrix also accounts for when the land cover changes. For example, if "Other" (e.g., vegetation) was truly present on Date 1 and was transformed into a "Building" on Date 2, then a hash mark would be placed in the OB diagonal cell as shown in Table 13-6. Conversely, if the "Other" pixel or polygon present on Date 1 was erroneously classified as "Road" instead of "Building" on the Date 2 map, then a hash mark would be placed in the off-diagonal OR cell of the matrix as shown. All hash marks placed in off-diagonal cells are change detection map classification errors. After the change detection error matrix is sufficiently populated with unbiased information obtained from the two classification maps, then it is possible to compute the overall accuracy, producer's accuracy, and user's accuracy as previously discussed.

The complete change transition error matrix (e.g., Table 13-6) provides the most useful information. However, it is only occasionally included in change studies because of the challenge of collecting sufficient data for so many categories especially when many of the categories are very rare occurrences. This makes simple random sampling inefficient. The problem can be improved somewhat using stratified sampling.

Assessing the Accuracy of a Binary Change Detection Map

Many change detection studies only attempt to determine if one class, e.g., forest, has changed from Date 1 to Date 2 (Im et al., 2008). This is called binary change detection (Im et al., 2009, 2011). Both the Date 1 and Date 2 remote sensing–derived thematic maps actually have two classes: forest and non-forest, i.e., a binary classification. When the two binary thematic maps are compared using post-classification comparison change detection (refer to Chapter 12) the result is a change detection map that can be analyzed using multiple-date ground reference information as shown in Table 13-7 (Khorram et al., 1999; Foody, 2002).

Foody (2010) suggests that this binary error matrix may be used to describe the sensitivity, specificity, and prevalence of the remote sensing–derived change map.

TABLE 13–7 A binary error matrix that can be used to compute the characteristics of a remote sensing–derived change detection map (based on Foody, G. M., 2010, "Assessing the Accuracy of Land Cover Change with Imperfect Ground Reference Data," *Remote Sensing of Environment*, 114:2271–2285).

		Multiple-date Binary Ground Reference Information		
		Change	No Change	Row total
Binary Map Class	Change	*a*	*b*	*g*
	No Change	*c*	*d*	*h*
	Column total	*e*	*f*	*N*

Sensitivity is the proportion of cases correctly classified as having changed (Foody, 2010):

$$\text{Sensitivity} = \frac{a}{a+c} = \frac{a}{e} \tag{13.9}$$

Specificity is the proportion of cases correctly predicted to have not changed:

$$\text{Specificity} = \frac{d}{b+d} = \frac{d}{f} \tag{13.10}$$

Sensitivity and specificity parameters represent the producer's accuracy for the change and no-change classes, respectively.

Evaluating the binary error matrix horizontally allows the predicted positive value (Foody, 2010):

$$\text{Predicted}_{pos} = \frac{a}{a+b} = \frac{a}{g} \tag{13.11}$$

and the predicted negative value:

$$\text{Predicted}_{neg} = \frac{d}{c+d} = \frac{d}{h} \tag{13.12}$$

to be determined. The positive and negative predicted values for the change detection map represent the user's accuracy (Lu and Weng, 2007; Liu et al., 2007) for the change and no-change classes, respectively.

Prevalence (θ) is computed using (Foody, 2010):

$$\text{Prevalence } (\theta) = \frac{a+c}{a+b+c+d} = \frac{e}{N} \tag{13.13}$$

Foody (2010) used these measures to compute the impact that error in ground reference data has on the accuracy of remote sensing–derived land-cover change detection maps.

Olofsson et al. (2014) identified additional good practices for assessing the accuracy and estimating the area of change in a land-cover change map. They provided recommendations for: 1) the sampling design, 2) the response design, and 3) the analysis that should take place.

Assessing the Accuracy of an Object-Based Image Analysis (OBIA) Classification Map

Lizarazo (2014) introduced a method to assess the accuracy of land-cover maps produced using object-based image analysis (OBIA). It is based on the use of classified "objects" in the map compared with reference "objects." Four STEP metrics are computed, including: 1) shape similarity (S); 2) theme similarity (T); 3) edge similarity (E); and 4) position similarity (P). Individual objects' similarity metrics are grouped by thematic class and expressed in the integrated STEP similarity matrix. The STEP indices and matrices can provide useful information about thematic and geometric accuracies of object-based image classifications. Moller et al. (2014) provide a framework for assessing the geometric accuracy of objects classified in an OBIA-derived classification map.

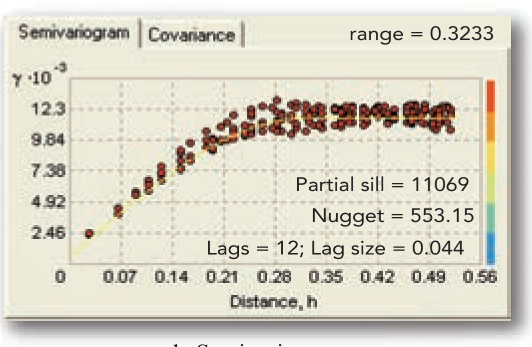
Geostatistical Analysis in Support of Accuracy Assessment

The emphasis in the previous sections on error evaluation is primarily concerned with the accuracy of *thematic* (categorical) information. However, the Earth's surface and remotely sensed images of that surface also have distinct *spatial* properties. Once quantified, these properties can be used for many tasks in remote sensing, including image classification and the sampling of both the image and the ground reference test data (Curran, 1988).

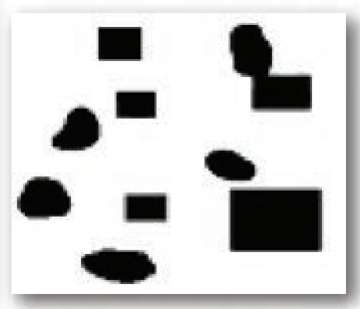
Where there is spatial dependence (i.e., autocorrelation) in a remote sensing–derived land-cover map, the number of sample points required to achieve a given level of confidence might be much fewer if the study is based on a systematic sample survey instead of a completely random survey (Curran, 1988). Constructing the empirical semivariogram model and evaluating its range and sill may make it possible to identify the critical autocorrelation characteristics of the dataset and design a sampling scheme that ensures that neighboring sample points are as far from one another as is practical for a fixed sample size and area (Van der Meer, 2012). This type of sampling scheme would minimize the duplication of information that often occurs in random sampling where some sample points are in-



a. Hypothetical crop polygons.



b. Semivariogram.



c. Predicted crop polygons.

FIGURE 13–7 a) A hypothetical map containing crop polygons. b) Semivariogram of the distribution of polygons. c) Predicted crop polygons. The semivariogram information can be used to improve accuracy assessment sampling design.

in random sampling where some sample points are inevitably close to one another. Curran (1988) describes how to test various grid-spacing designs that allow sampling error and sample size to be considered in a spatial context. He suggests that a well-designed systematic survey based on geostatistical analysis can increase the precision over random sampling and decrease the sometimes impossibly large sample sizes associated with random sampling (Curran and Williamson, 1986).

For example, consider the hypothetical map of remote sensing–derived agricultural crop polygons shown in Figure 13-7. Its semivariogram (discussed in Chapter 4) and predicted image are also provided. The geostatistical spherical kriging process did a good job of predicting the original surface. Plus, we now have detailed information about the agricultural field autocorrelation present in the dataset. If this type of information were available over an entire region, it would be possible to identify the optimum sampling frequency that should be used for a) identifying the optimum spatial resolution of the remote sensing system, b) the spacing of individual frames of imagery if the study involves a sampling scheme rather than wall-to-wall mapping and/or c) allocating ground reference test sites to perform a classification accuracy assessment.

Jacobs and Thomas (2003) used fuzzy set analysis and geospatial statistics (kriging) to create a spatial view of the accuracy of an Arizona GAP Analysis map. Kyriakidis et al. (2005) described geostatistical approaches for mapping thematic classification accuracy. Zhu and Stein (2006) and Van der Meer (2012) review how geostatistics can be used to design sampling schemes and, prior to actually sampling these in the field, determine the kriging estimation variance that will be obtained. Burnicki (2011) used geostatistical analysis to examine the spatial and temporal patterns of error in time-series individual maps and the change detection map.



Image Metadata and Lineage Information for Remote Sensing–Derived Products

It is becoming increasingly important to document a) all of the information about the creation of an individual remotely sensed image, and b) the various procedures and processes applied to the imagery to create an output product and assess its accuracy.

Individual Image Metadata

The U.S. government now requires that all imagery collected for government use contain detailed metadata (data about data) as to how the images were collected. Detailed metadata standards have been developed for raster image datasets by the U.S. Federal Geographic Data Committee (FGDC). This type of information allows the image analyst and decision-maker to document the source of the imagery and all of its inherent characteristics. Metadata standards are an essential component of national and global spatial data infrastructures (SDI) (Jensen et al., 2002).

Lineage of Remote Sensing–Derived Products

It is not sufficient to collect only image metadata. It is usually necessary to carefully document every operation that is performed on the remote sensing data that results in an output product such as a land-use or land-cover map. This is commonly referred to as *lineage documentation*. Unfortunately, manual bookkeeping of the processes used to create a final product is cumbersome and rarely performed. Some digital image processing systems do provide history or audit files to keep track of the iterations and operations performed. However, none of these methods is capable of fulfilling the information requirements of a true lineage report that itemizes the characteristics of both the image and

cartographic sources, the topological relationships among sources, intermediate and final product layers, and a history of the transformations applied to the sources to derive the output products (Lanter, 1991).

Lineage information should be included in *quality assurance reports* associated with every remote sensing–derived thematic product, including:

- a summary of all image and ancillary source materials (e.g., soil and geology maps, digital terrain models) used to derive the thematic information;
- geoid, datum, and map projection information used to create the thematic information;
- geometric rectification and radiometric correction parameters;
- the processing steps used to transform the remotely sensed data into information;
- the methods and results of the accuracy assessment;
- the location of the archived original, interim, and final datasets;
- procedures for accessing the archived information; and
- geometric and thematic attribute accuracy assessment, if available.

Quality assurance is an important part of life today. Image analysts who extract thematic information from remotely sensed data add value and rigor to the product by documenting its lineage.



References

Anderson, J. R., Hardy, E., Roach, J., and R. Witmer, 1976, *A Land-Use and Land-Cover Classification System for Use with Remote Sensor Data*, Washington: U.S. Geological Survey, Professional Paper #964, 28 p.

Atkinson, P., and P. Curran, 1995, "Defining an Optimal Size of Support for Remote Sensing Investigations," *IEEE Transactions Geoscience Remote Sensing*, 33:768–776.

Borengasser, M., Hungate, W. S., and R. Watkins, 2008, *Hyperspectral Remote Sensing: Principles and Application*s, Boca Raton: CRC Press, 119 p.

Bossler, J. D., Jensen, J. R., McMaster, R. B., and C. Rizos, 2002, *Manual of Geospatial Science and Technology*, London: Taylor & Francis, 623 p.

Burnicki, A., 2011, "Spatio-temporal Errors in Land-Cover Change Analysis: Implications for Accuracy Assessment," *Intl. Journal of Remote Sensing*, 32(22):7487–7512.

Comber, A., Fisher, P., Brunsdon, C., and A. Khmag, 2012, "Spatial Analysis of Remote Sensing Image Classification Accuracy," *Remote Sensing of Environmen*t, 127:237–246.

Congalton, R. G., 1981, *The Use of Discrete Multivariate Analysis for the Assessment of Landsat Classification Accuracy*, Blacksburg: Virginia Polytechnic Institute and State University, Master's thesis.

Congalton, R. G., 1988, "Using Spatial Autocorrelation Analysis to Explore the Errors in Maps Generated from Remotely Sensed Data," *Photogrammetric Engineering & Remote Sensing*, 54(5):587–592.

Congalton, R. G., 1991, "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data," *Remote Sensing of Environment*, 37:35–46.

Congalton, R. G., 2005, "Chapter 1: Putting the Map Back in Accuracy Assessment," in Lunetta, R. L. and J. G. Lyon, (Eds.), *Remote Sensing and GIS Accuracy Assessment*, Boca Raton: CRC Press, 1–14 p.

Congalton, R. G., and K. Green, 2009, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd Ed., Boca Raton: CRC Press, 183 p.

Congalton, R. G., and R. A. Mead, 1983, "A Quantitative Method to Test for Consistency and Correctness in Photointerpretation," *Photogrammetric Engineering & Remote Sensing*, 49(1):69–74.

Congalton, R. G., Oderwald, R. G., and R. A. Mead, 1983, "Assessing Landsat Classification Accuracy Using Discrete Multivariate Statistical Techniques," *Photogrammetric Engineering & Remote Sensing*, 49(12):1671–1678.

Curran, P. J., 1988, "The Semivariogram in Remote Sensing," *Remote Sensing of Environment*, 24:493–507.

Curran, P. J., and H. D. Williamson, 1986, "Sample Size for Ground and Remotely Sensed Data," *Remote Sensing of Environment*, 20:31–41.

Fitzpatrick-Lins, K., 1981, "Comparison of Sampling Procedures and Data Analysis for a Land-use and Land-cover Map," *Photogrammetric Engineering & Remote Sensing*, 47(3):343–351.

Foody, G. M., 2002, "Status of Land Cover Classification Accuracy Assessment," *Remote Sensing of Environment*, 80:185–201.

Foody, G. M., 2005, "Local Characterization of Thematic Classification Accuracy through Spatially Constrained Confusion Matrices," *International Journal of Remote Sensing*, 26:1217–1228.

Foody, G. M., 2008, "Harshness in Image Classification Accuracy Assessment," *International Journal of Remote Sensing*, 29(11):3137–3158.

Foody, G. M., 2009, "Sample Size Determination for Image Classification Accuracy Assessment and Comparison," *International Journal of Remote Sensing*, 30(20):5273–5291.

Foody, G. M., 2010, "Assessing the Accuracy of Land Cover Change with Imperfect Ground Reference Data," *Remote Sensing of Environment*, 114:2271–2285.

Gomez, D., Bigng, G., and J. Montero, 2008, "Accuracy Statistics for Judging Soft Classification," *International Journal of Remote Sensing*, 29(3):693–709.

Gopal, S., and C. Woodcock, 1994, "Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets," *Photogrammetric Engineering & Remote Sensing*, 60(2):181–188.

Green, K., and R. G. Congalton, 2003, Chapter 12: "An Error Matrix Approach to Fuzzy Accuracy Assessment: The NIMA Geocover Project," in Lunetta, R. and J. Lyon, (Eds.), *Geospatial Data Accuracy Assessment,* EPA Report #600/R-03/064 (December, 2003; 339 p.), Washington: U.S. Environmental Projection Agency, 191–200.

Green, K., and R. G. Congalton, 2005, "An Error Matrix Approach to Fuzzy Accuracy Assessment—The NIMA Geocover Project," in Lunetta, R., and J. Lyon (Eds.), *Remote Sensing and GIS Accuracy Assessment*, Boca Raton: CRC Press, 204–215.

Im, J., Rhee, J., and J. R. Jensen, 2009, "Enhancing Binary Change Detection Performance Using a Moving Threshold Window (MTW) Approach," *Photogrammetric Engineering & Remote Sensing*, 75(8):951–962.

Im, J., Rhee, J., Jensen, J. R., and M. E. Hodgson, 2007, "An Automated Binary Change Detection Model Using a Calibration Approach," *Remote Sensing of Environment*, 106:89–105.

Im, J., Lu, Z., and J. R. Jensen, 2011, "A Genetic Algorithm Approach to Moving Threshold Optimization for Binary Change Detection," *Photogrammetric Engineering & Remote Sensing*, 77(2):167–180.

Im, J., Rhee, J., Jensen, J. R., and M. E. Hodgson, 2008, "Optimizing the Binary Discriminant Function in Change Detection Applications," *Remote Sensing of Environment*, 112:2761–2776.

Jacobs, S. R., and K. A. Thomas, 2003, "Fuzzy Set and Spatial Analysis Techniques for Evaluating the Thematic Accuracy of a Land-Cover Map," in Lunetta, R. L. and J. G. Lyons (Eds.), *Geospatial Data Accuracy Assessment,* Las Vegas: EPA, Report No. EPA/600/R-03/064, 335 p.

Jensen, J. R., and J. Im, 2007, "Remote Sensing Change Detection in Urban Environments," in R.R. Jensen, J. D. Gatrell and D. D. McLean (Eds.), *Geo-Spatial Technologies in Urban Environments Policy, Practice, and Pixels*, (2nd Ed.), Berlin: Springer-Verlag, 7–32.

Jensen, J. R., and R. R. Jensen, 2013, *Introductory Geographic Information Systems*, Boston: Pearson, 400 p.

Jensen, J. R., Botchway, K., Brennan-Galvin, E., Johannsen, C., Juma, C., Mabogunje, A., Miller, R., Price, K., Reining, P., Skole, D., Stancioff, A., and D. R. F. Taylor, 2002, *Down to Earth: Geographic Information for Sustainable Development in Africa*, Washington: National Research Council, 155 p.

Jensen, J. R., Cowen, D. J., Narumalani, S., Althausen, J. D., and O. Weatherbee, 1993, "An Evaluation of CoastWatch Change Detection Protocol in South Carolina," *Photogrammetric Engineering & Remote Sensing*, 59:1039– 1046.

Jensen, J. R., Guptill, S., and D. Cowen, 2012, *Change Detection Technology Evaluation*, Bethesda: U.S. Bureau of the Census, Task 2007, FY2012 Report, 232 p.

Jensen, J. R., Hadley, B. C., Tullis, J. A., Gladden, J., Nelson, S., Riley, S., Filippi, T., and M. Pendergast, 2003, *2002 Hyperspectral Analysis of Hazardous Waste Sites on the Savannah River Site*, Aiken, SC: Westinghouse Savannah River Company, WSRC-TR-2003-0025, 52 p.

Jensen, J. R., Im, J., Hardin, P., and R. R. Jensen, 2009, "Chapter 19: Image Classification," in *The Sage Handbook of Remote Sensing*, Warner, T. A., Nellis, M. D. and G. M. Foody (Eds.), 269–296.

Jensen, R. R., and J. M. Shumway, 2010, "Sampling Our World," in B. Gomez and J. P. Jones III (Eds.), *Research Methods in Geography*, New York: Wiley Blackwell, 77–90.

Khorram, S., Biging, G. S., Chrisman, N. R., Colby, D. R., Congalton, R.G., Dobson, J. E., Ferguson, R. L., Goodchild, M. F., Jensen, J. R., and T. H. Mace, 1999, *Accuracy Assessment of Remote Sensing-derived Change Detection*, Bethesda: American Society for Photogrammetry & Remote Sensing, 78 p.

Kyriakidis, P. C., Liu, X., and M. F. Goodchild, 2005, "Geostatistical Mapping of Thematic Classification Uncertainty," in Lunetta, R. L. and J. G. Lyon (Eds.), *Remote Sensing and GIS Accuracy Assessment*, Boca Raton: CRC Press, 184–203.

Lanter, D. P., 1991, "Design of a Lineage-based Meta-database for GIS," *Cartography and Geographic Information Systems*, 18(4):255–261.

Liu, C., Frazier, P., and L. Kumar, 2007, "Comparative Assessment of the Measures of Thematic Classification Accuracy," *Remote Sensing of Environment*, 107:606–616.

Lizarazo, I., 2014, "Accuracy Assessment of Object-based Image Classification: Another STEP," *International Journal of Remote Sensing*, 35(16):6135–6156.

Lu, D., and Q. Weng, 2007, "A Survey of Image Classification Methods and Techniques for Improving Classification Performance," *International Journal of Remote Sensing*, 28, 823–870.

Lunetta, R. L., and J. G. Lyon (Eds.), 2003, *Geospatial Data Accuracy Assessment,* Las Vegas: EPA, Report No. EPA/600/R-03/064.

Lunetta, R. L., and J. G. Lyon, (Eds.), 2005, *Remote Sensing and GIS Accuracy Assessment*, Boca Raton: CRC Press, 380 p.

Lunetta, R. S., Congalton, R. G., Fenstermaker, L. K., Jensen, J. R., McGwire, K. C., and L. R. Tinney, 1991, "Remote Sensing and Geographic Information System Data

Integration: Error Sources and Research Issues," *Photogrammetric Engineering & Remote Sensing,* 57(6):677–687.

Macleod, R. D., and R. G. Congalton, 1998, "A Quantitative Comparison of Change-Detection Algorithms for Monitoring Eelgrass from Remotely Sensed Data," *Photogrammetric Engineering & Remote Sensing*, 64(3):207–216.

McCoy, R., 2005, *Field Methods in Remote Sensing*, NY: Guilford, 159 p.

Meyer, M., and L. Werth, 1990, "Satellite Data: Management Panacea or Potential Problem?" *Journal of Forestry*, 88(9):10–13.

Moller, M., Birger, J., Gidudu, A., and C. Glaber, 2014, "A Framework for the Geometric Accuracy Assessment of Classified Objects," *International Journal of Remote Sensing*, 34(24):8685–8698.

Morisette, J. T., Privette, J. L., Strahler, A., Mayaux, P., and C. O. Justice, 2005, "Validation of Global Land-Cover Products by the Committee on Earth Observing Satellites," in Lunetta, R. L. and J. G. Lyon (Eds.), *Remote Sensing and GIS Accuracy Assessment*, Boca Raton: CRC Press, 36–46.

Muchoney, D. M., and A. H. Strahler, 2002, "Pixel- and Site-based Calibration and Validation Methods for Evaluating Supervised Classification of Remotely Sensed Data," *Remote Sensing of Environment*, 81:290–299.

NRC, 2007, *Contributions of Land Remote Sensing for Decisions About Food Security and Human Health—Workshop Report*, Washington: National Academy Press, 230 p.

NRC, 2009, *Uncertainty Management in Remote Sensing of Climate Date*, Washington: National Academy Press, 64 p.

NRC, 2012, *Ecosystem Services—Charting A Path to Sustainability*, Washington: National Academy Press, 121 p.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and M. A. Wulder, 2014, "Good Practices for Estimating Area and Assessing Accuracy of Land Change," *Remote Sensing of Environment*, 148:42–57.

Paine, D. P., and J. D. Kiser, 2003, "Chapter 23: Mapping Accuracy Assessment," *Aerial Photography and Image Interpretation*, 2nd Ed., NY: John Wiley & Sons, 465–480.

Pontius, R. G., and M. L. Cheuk, 2006, "A Generalized Cross-tabulation Matrix to Compare Soft-classified maps at Multiple Spatial Resolutions," *International Journal of Geographical Information Science*, 20:1–30.

Pontius, R. G., and M. Millones, 2011, "Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment," *International Journal of Remote Sensing*, 32(15):4407–4429.

Pouliot, D., Latifovic, R., Zabcic, N., Guidon, L., and I. Olthof, 2014, "Development and Assessment of a 250 m Spatial Resolution MODIS Annual Land Cover Time Series (2000–2011) for the Forest Region of Canada Derived from Change-based Updating," *Remote Sensing of Environment*, 140:731–743.

Stehman, S. V., 1997, "Selecting and Interpreting Measures of Thematic Classification Accuracy," *Remote Sensing of Environment*, 62:77–89.

Stehman, S. V., 2000, "Practical Implications of Design-based Sampling for Thematic Map Accuracy Assessment," *Remote Sensing of Environment*, 72:35–45.

Stehman, S. V., 2001, "Statistical Rigor and Practical Utility in Thematic Map Accuracy Assessment," *Photogrammetric Engineering & Remote Sensing*, 67:727–734.

Stehman, S. V., 2009, "Sampling Designs for Accuracy Assessment of Land Cover," *International Journal of Remote Sensing*, 30(20):5243–5272.

Stehman, S. V., and G. M. Foody, 2009, "Accuracy Assessment," in Warner, T. A., Nelllis, M. D. and G. M. Foody (Eds.), *The Sage Handbook of Remote Sensing*, Los Angeles: Sage Publications, 297–309.

Stehman, S. V., and J. D. Wickham, 2011, "Pixels, Blocks of Pixels, and Polygons: Choosing A Spatial Unit for Thematic Accuracy Assessment," *Remote Sensing of Environment*, 115:3044–3055.

Stehman, S. V., and R. L. Czaplewski, 1998, "Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles," *Remote Sensing of Environment*, 64:331–344.

Stehman, S. V., Olofsson, P., Woodcock, C. E., Herold, M., and M. A. Friedl, 2012, "A Global Land-cover Validation Data Set II: Augmenting a Stratified Sampling Design to Estimate Accuracy by Region and Land-cover Class," *International Journal of Remote Sensing*, 33(22):6975–6993.

Story, M., and R. Congalton, 1986, "Accuracy Assessment: A User's Perspective," *Photogrammetric Engineering & Remote Sensing*, 52(3):397–399.

Van der Meer, F., 2012, "Remote-sensing Image Analysis and Geostatistics," *International Journal of Remote Sensing,* 33(18):5644–5676.

van Oort, P. A. J., 2007, "Interpreting the Change Detection Error Matrix," *Remote Sensing of Environment*, 108:1–8.

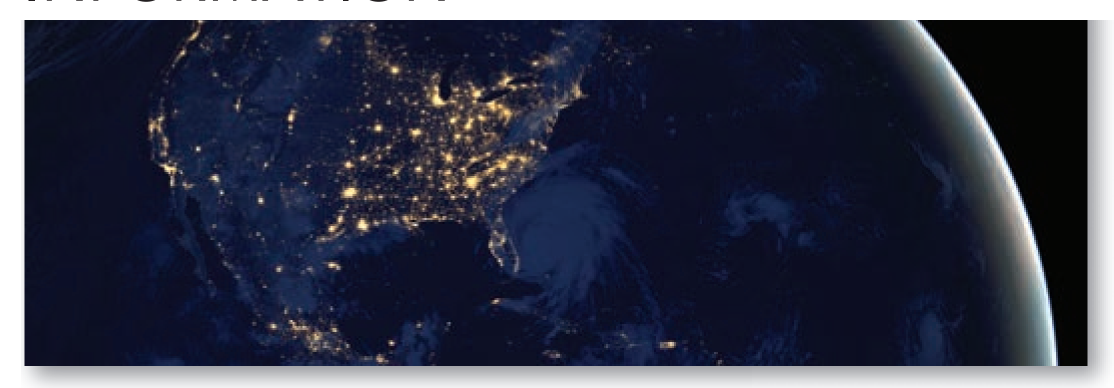
Warner, T. A., Almutairi, A., and J. Y. Lee, 2009, "Remote Sensing of Land Cover Change," in Warner, T. A., Nelllis, M. D. and G. M. Foody (Eds.), *The Sage Handbook of Remote Sensing*, Los Angeles: Sage Publications, 459–472.

Wickham, J. D., Stehman, S. V., Fry, J. A., Smith, J. H., and C. G. Homer, 2010, "Thematic Accuracy of the NLCD 2001 Land Cover for the Conterminous United States," *Remote Sensing of Environment*, 114:1286–1296.

Wickham, J. D., Stehman, S. V., Gass, L., Dewitz, J., Fry, J. A., and T. G. Wade, 2013, "Accuracy Assessment of NLCD 2006 Land Cover and Impervious Surface," *Remote Sensing of Environment*, 130:294–305.

Zhu, Z., and M. Stein, 2006, "Spatial Sampling Design for Prediction with Estimated Parameters," *Journal of Agricultural, Biological, and Environmental Statistics*, 11:24–44.

APPENDIX: SOURCES OF IMAGERY AND OTHER GEOSPATIAL INFORMATION



Source: NOAA

Obtaining timely, accurate, geospatial information is one of the most important and challenging activities of remote sensing and GIS investigations. Chapter 2 reviewed numerous types of remote sensing data acquired using airborne and satellite remote sensing systems. This Appendix provides more detailed information about how to access selected types of remote sensing and other types of geospatial data using the Internet. The datasets listed are not exhaustive, but will hopefully provide sufficient information to evaluate and obtain some of the most commonly used types of remote sensing and other types of geospatial data.

This Appendix begins by introducing four federal geospatial data search engines and image repositories. Next, three representative commercial geospatial data search engines and/or repositories are reviewed. The remainder of the Appendix is devoted to a discussion of selected: 1) thematic datasets (e.g., elevation, hydrology, land use/land cover), 2) sources of public remote sensor data, and 3) sources of commercial and international remote sensor data.



Table of Contents

USGS *EarthExplorer* User Interface

Geographic Search Area

a.

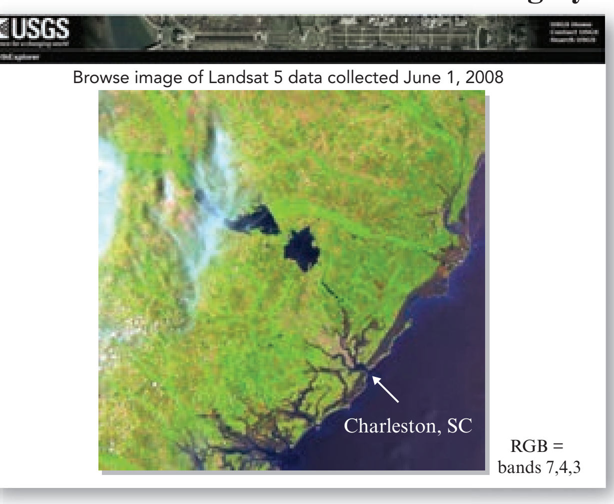


FIGURE A–1 a) The USGS *EarthExplorer* user interface (http://earthexplorer.usgs.gov/). The geographic search area is available online. A search for Landsat 4 and 5 imagery of Charleston, SC, from 2005 to 2010 in June was selected. b) This June 1, 2008, Landsat 5 image was one of several images that met the search criteria (imagery courtesy of NASA).

Land Use/Land Cover and Biodiversity/Habitat Data

- NLCD—National Land Cover Dataset (USGS)
- C-CAP—Coastal Change Analysis Program (NOAA)
- GAP Analysis Program (USGS)
- NWI—National Wetlands Inventory (USFWS)
- NPN—U.S. National Phenology Network

Population Demographic Data

- MAF/TIGER—Master Address File/Topologically Integrated Geographic Encoding and Reference System (Bureau of the Census)
- 2010 Census Demographic Data (Bureau of the Census)
- LandScan (Oak Ridge National Laboratory)

Remote Sensor Data—Public

- Selected publicly available sources of analog and digital remote sensor data.

Remote Sensor Data—Commercial and International

- Selected commercial and international sources of analog and digital remote sensor data.



Federal Geospatial Data Search Engines and Repositories

A powerful characteristic of federal geospatial data repositories is that the data are in the public domain and available to the general public at no expense. Only a select few of the many federal image and geospatial repositories are introduced here.

USGS *EarthExplorer*

One of the best places to look for remote sensing data and various remote sensing–derived products is at the U.S. Geological Survey's *EarthExplorer* website (http://earthexplorer.usgs.gov/). This site allows users to select a geographic place or region by typing in the name of a location, entering geographic coordinates, or drawing on a map (Figure A-1a). Once the geographic area has been identified, the user can identify specific types of data from the menu. This includes access to digitized aerial photography (discussed below), AVHRR, Calibration/Validation Reference Sites, Commercial (IKONOS-2), Declassified data, Digital Elevation (ASTER Global DEM, GMTED2010, GTOPO30, GTOPO30 HYDRO 1K, SRTM, SRTM Void Filled), Digital Line Graphs, EO-1 (ALI, Hyperion), Forest Carbon Sites, Landsat (Global Land Survey, Archive, CDR, Legacy, MRLC), LiDAR, NASA SPDAAC Collections, Orbview-3, RADAR (SIR-C), Vegetation monitoring, HCMM, and Land Cover (NLCD 1992, 2001, 2006). For example, the user interface in Figure A-1a was used to search for Landsat 5 Thematic Mapper imagery of

The National Map

FIGURE A–2 *The National Map* user interface is especially useful for locating information about geographic boundaries, terrain elevation, geographic names, hydrography, land cover, orthoimagery, structures, and transportation (http://nationalmap.gov/viewer.html). This is a display of 1 × 1 ft. orthoimagery of Charleston, SC (aerial photography courtesy of the U.S. Geological Survey).

Charleston, SC, collected in June from 2005 to 2010. One of the results of the search is shown in Figure A-1b. The user may discard the search or order a digital copy of the June 2, 2008, Landsat image.

From 1972 to 2005, the USGS Earth Resources Observation and Science (EROS) Center provided remote sensing *film*-based products to the public. EROS is home to an archive of 12 million frames of analog photography ranging from 1937 to the present. The archive contains analog collections from both aerial and satellite platforms including programs such as the National High Altitude Program (NHAP), National Aerial Photography Program (NAPP), U.S. Antarctic Resource Center (USARC), Declassified 1 (CARONA, ARGON, and LANYARD), Declassified 2 (KH-7 and KH-9), and Landsat (1972 – 1992, Landsat 1-5). Since 2004, EROS has been digitally scanning the analog collection at resolutions ranging from 64 μm (400 dpi) to 25 μm (1,000 dpi) depending upon the severity of the *vinegar* syndrome (a slow form of chemical deterioration that causes the film to shrink, buckle, and become unusable). The digitized imagery is available at no cost through the *EarthExplorer* interface (Moe and Longhenry, 2013).

Users must register with *EarthExplorer* to save searches and order data. Some digital datasets can be downloaded immediately. Other datasets take a few hours or a couple of days for the data to be extracted and then made available to the user via FTP. Normally, users are not sent to additional sites to order data. This is an advanced search engine that is very useful for those who know the type and characteristics of the imagery and/or other geospatial data they are looking for.

USGS *The National Map*

Managed by the National Geospatial Program (NGP), *The National Map* viewer (Figure A-2) provides data visualization and download of all eight *National Map* datasets, including: geographic boundaries, elevation, geographic names, hydrography, land cover, orthoimagery, structures, transportation, and US Topo products (Carswell, 2011; USGS NGP, 2014). The majority of *The National Map* effort is devoted to acquiring and integrating medium-scale (nominally 1:24,000-scale) geospatial data for the eight base layers from a variety of sources and providing access to the resulting seamless coverages of geospatial data.

An orthoimage is remotely sensed image data in which displacement of features in the image caused by terrain

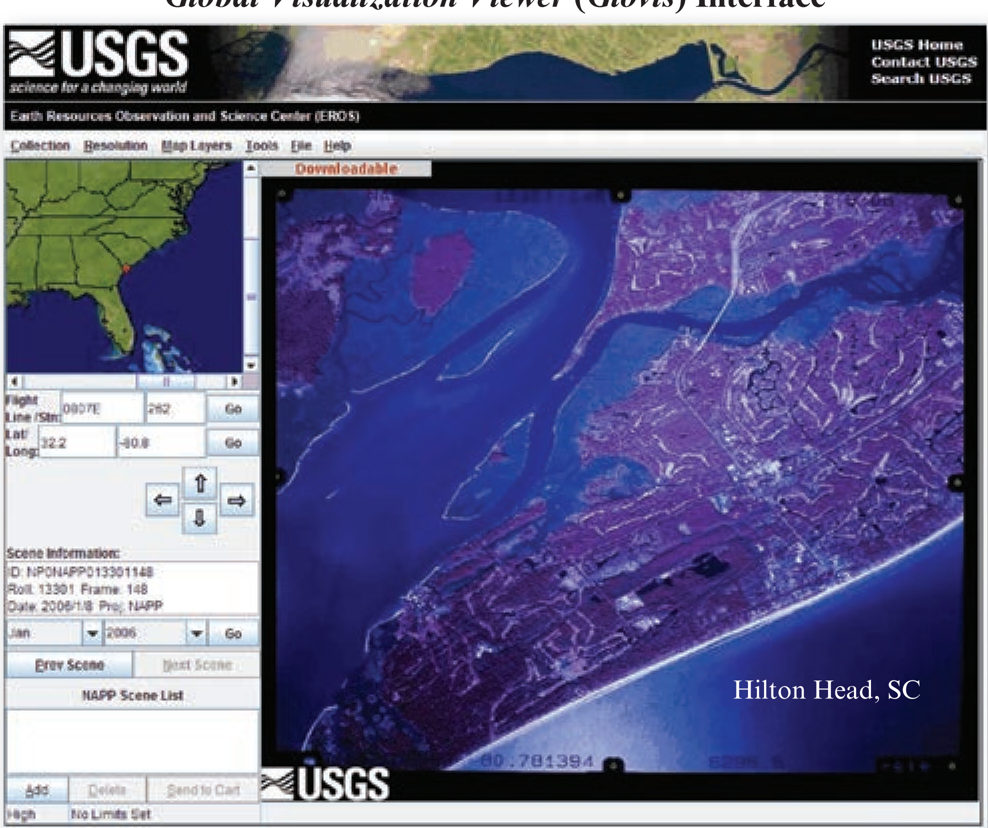


FIGURE A–3 The *Global Visualization Viewer (Glovis)* user interface provides information about the availability of the following types of imagery: aerial (NHAP, NAPP), ASTER, EO-1 (Advanced Land Imager, Hyperion), Landsat archive, Landsat Global Land Survey, Landsat MRLC collections, Landsat Legacy Collections, MODIS (Aqua, Terra, combined), and Terra Outlook (http://glovis.usgs.gov/). This is a display of high spatial resolution NAPP aerial photography of Hilton Head, SC, collected on January 8, 2006. The image can be downloaded directly from the *Glovis* website (aerial photography courtesy of U.S. Geological Survey).

relief and sensor orientation have been mathematically removed. Orthoimagery combines the image characteristics of a photograph with the geometric spatial accuracy and reliability of a planimetric map. USGS digital orthoimage resolution may vary from 6 in. to 1 m. *The National Map* provides free downloads of public domain, 1 × 1 m orthoimagery for the conterminous U.S. with many urban areas and other locations at 2-ft. or finer resolution. For example, 1 × 1 ft. spatial resolution orthoimagery of Charleston, SC, in *The National Map* database is displayed in Figure A-2.

The National Map viewer also allows visualization and identification queries (but not downloads) of other data, including: Ecosystems, Protected Areas, Gap Analysis Program Land Cover, Hazards, Weather, Wetlands, Public Land Survey System, and National Park Service Boundaries. To research and download historical USGS data, such as Digital Orthophoto Quadrangles and Digital Line Graphs, or to access specific LiDAR point cloud data, use the USGS *EarthExplorer* website.

USGS *Global Visualization Viewer*

The USGS *Global Visualization Viewer* (*Glovis*) is a straightforward tool for searching and ordering specific types of satellite and aerial imagery. The types of data that can be searched include aerial (NHAP, NAPP), ASTER, EO-1 (Advanced Land Imager, Hyperion), Landsat archive, Landsat Global Land Survey, Landsat MRLC collections, Landsat Legacy Collections, MODIS (Aqua, Terra, combined), and Terra Outlook.

A search for National Aerial Photography Program (NAPP) imagery of Hilton Head, SC, is shown in Figure A-3. The color-infrared aerial photograph was obtained on January 8, 2006. If desired, this image can be downloaded directly from the Glovis interface.

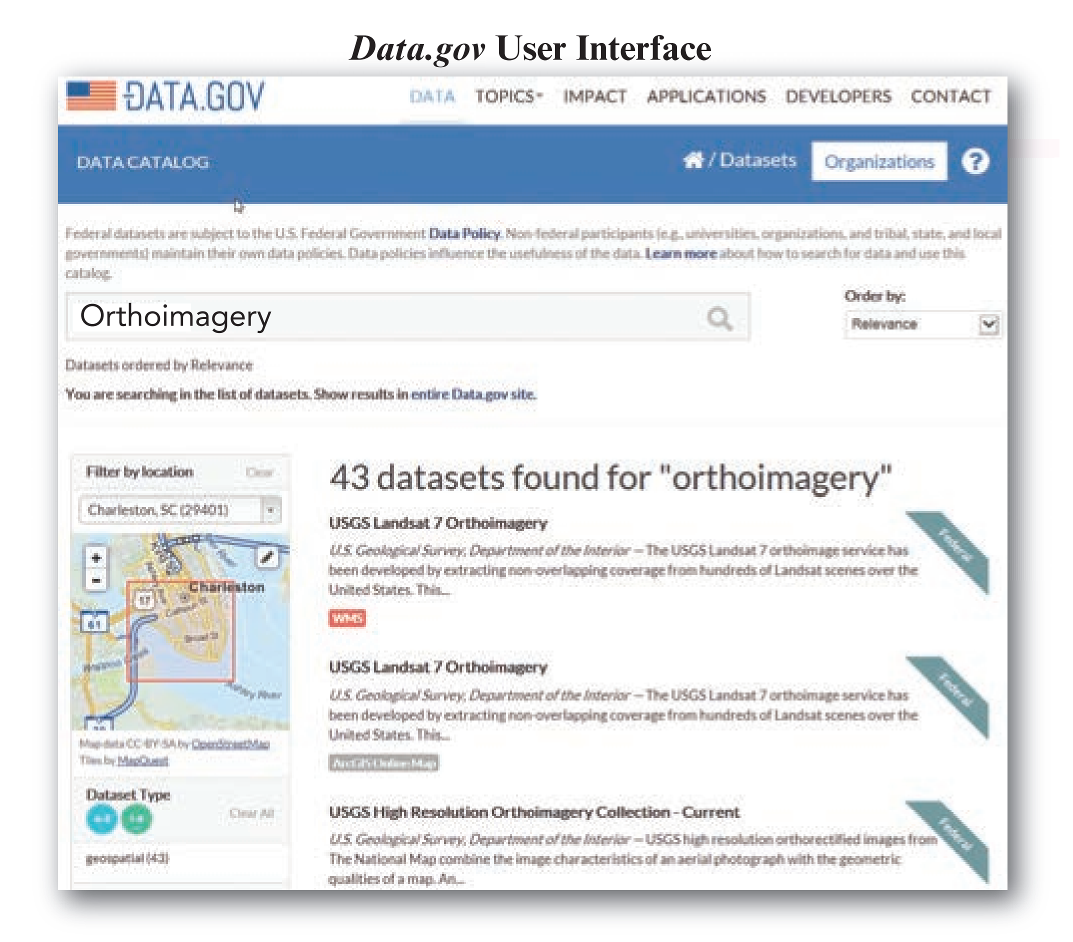
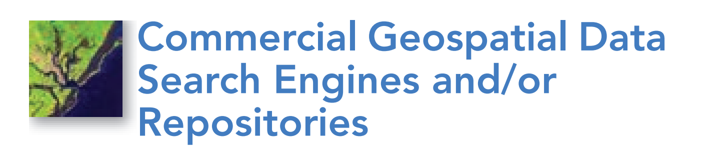


FIGURE A–4 The *Data.gov* user interface (http://catalog.data.gov/dataset) allows the user to filter by location and then search a list of datasets by type. In this example, the user specified the "geospatial dataset type" and requested information on all the "Orthoimagery" for Charleston, SC. The results of the search identified 43 orthoimagery datasets available for the study area. The user can evaluate the metadata associated with each file and/or download the data (courtesy of Data.gov).

Data.gov

On October 1, 2011, *Data.gov* was introduced as "the home of the U.S. government's open data." With this move, U.S. national geospatial assets were brought together in one place to theoretically make it easier for the public to browse and access over 400,000 maps, datasets, and geospatial services (Figure A-4). In addition, the services, mapping and visualization capabilities, and data standards behind these sites also became accessible. This work was done in coordination with a refresh of the Geospatial Platform, which remains the online home for the Federal Geographic Data Committee's guidance, policies, and standards.

Data.gov allows the user to search for geospatial information by searching for a particular type of dataset in a particular geographic location (Figure A-4). In this example, the location was specified as Charleston, SC, and the geospatial database was searched to find "Orthoimagery." The result was a list of 43 orthoimagery datasets. This website is ideal for less experienced users who are not absolutely sure what image or other geospatial data they are looking for.



Commercial Geospatial Data Search Engines and/or Repositories

There are numerous commercial image and geospatial search engines such as *Google* earth, bing, *ArcGIS Online*, and others. Selected characteristics of these three geospatial search engines are introduced for demonstration purposes.

Google, Inc., *Google* earth Search Engine

The *Google* earth website provides detailed image and other geospatial information about most locations in the world. The user can specify the latitude, longitude coordinates or interactively locate the area of interest on the synthetic *Google* earth three-dimensional globe.

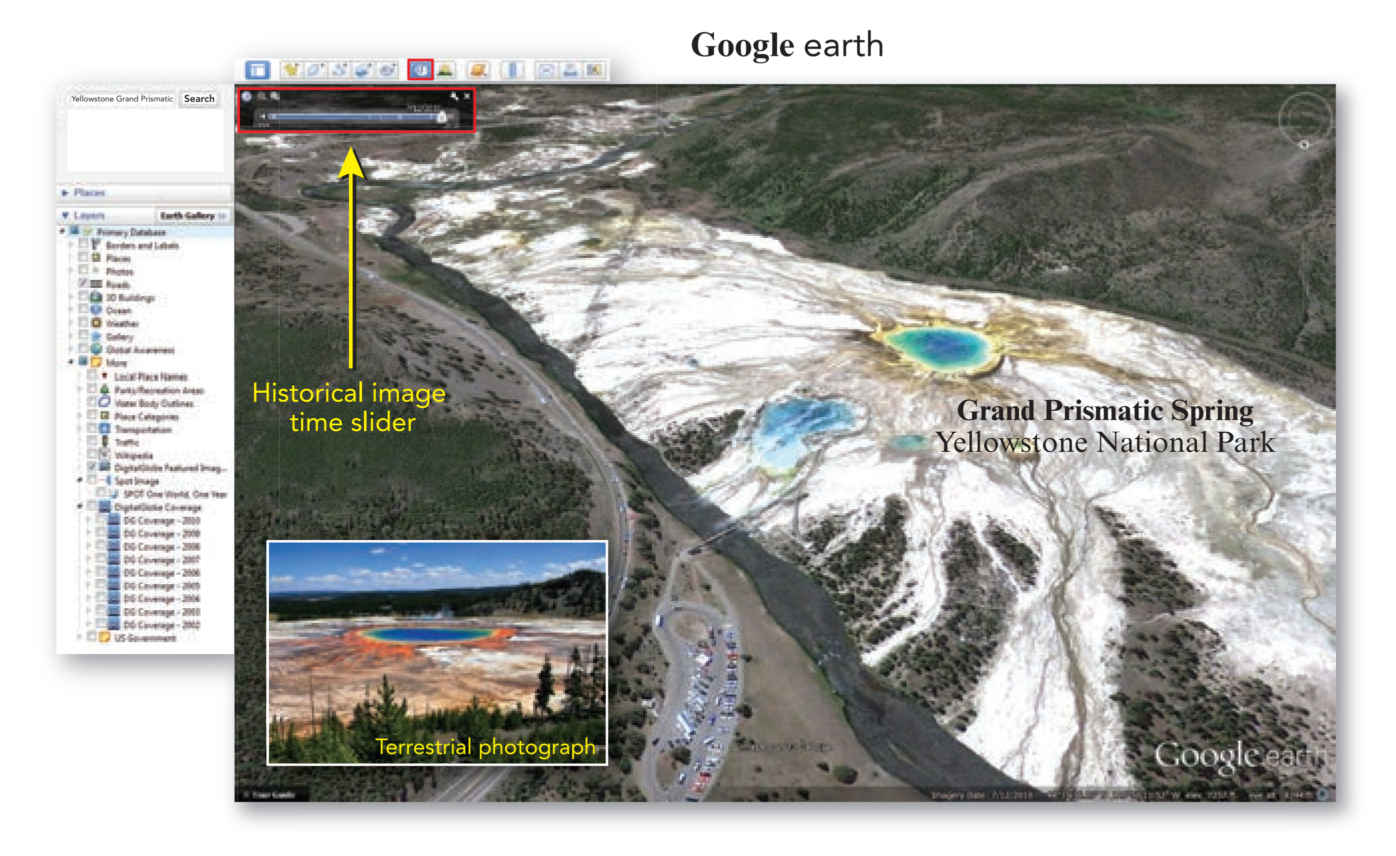


FIGURE A–5 *Google* earth can be used to look at imagery of almost anywhere on the Earth. In this example, the user requested geospatial information about the Grand Prismatic Spring in Yellowstone National Park in Wyoming. The historical image slider can be used to select for display any of the remote sensing imagery stored in the database. In this particular case, a DigitalGlobe GeoEye-1 natural-color image of the Grand Prismatic Spring obtained on July 12, 2010, is displayed (user interface courtesy of *Google* earth, Inc.; imagery courtesy of DigitalGlobe, Inc.; terrestrial photograph provided by the author).

Google earth relies heavily on imagery acquired by DigitalGlobe, Inc. (e.g., IKONOS, GeoEye, QuickBird, and WorldView-1 and -2). For example, Figure A-5 depicts one image derived for a search on the Grand Prismatic Spring in Yellowstone National Park in Wyoming. The July 12, 2010, GeoEye-1 natural color image is draped over U.S. national elevation data (NED) yielding a very informative three-dimensional display of the spring and surrounding roads and mountainous terrain. *Google* earth also uses public-domain imagery such as that provided by the U.S. Department of Agriculture Farm Service Agency.

The simple historical image slider user-interface in the upper left-hand corner of the interface can be used to interactively display each of the dates of imagery of the study area that reside in the database. In this particular case the following images are available:

- 8/24/1994 USGS Black and White
- 10/23/2003 DigitalGlobe
- 11/6/2004 DigitalGlobe
- 8/10/2006 DigitalGlobe
- 9/29/2006 USDA Farm Service Agency
- 8/27/2009 USDA Farm Service Agency
- 7/12/2010 GeoEye-1

In the Layer sidebar, *Google* earth provides information about all the DigitalGlobe and SPOT image data available for the study area as well as other thematic information such as roads that can be overlaid on the imagery. Users often extract screen captures of the imagery displayed using *Google* earth for a variety of non-scientific purposes. It is necessary to contact the original data provider such as DigitalGlobe, Inc. or SPOT Image, Inc. to obtain the original remote sensor data.

Microsoft, Inc., bing Search Engine

The Microsoft, Inc., bing search engine provides detailed image information about most locations in the world. The user can specify the latitude, longitude coordinates or interactively locate the area of interest using the bing interface. bing relies heavily on digital frame camera aerial photography (e.g., UltraCam) acquired by photogrammetric engineering firms such as

Microsoft, Inc. bing

FIGURE A–6 The Microsoft Inc., bing interface was used to view imagery of downtown New York City, NY. High spatial resolution 3Di digital aerial photography is displayed. The user has also specified that Bird's Eye three-dimensional building information available from Pictometry International, Inc. be displayed on top of the remote sensor data (image courtesy of Pictometry International, Inc.; user interface courtesy of Bing, Microsoft, Inc.).

3Di, Sanborn Map Company, Pictometry International, Vexcel Imaging, and others. For example, Figure A-6 depicts high spatial resolution natural color digital aerial photography of a search centered on downtown New York City, NY. Three-dimensional building information provided by Pictometry International's Bird's Eye view is overlaid on top of the imagery. bing also uses public-domain and satellite imagery. Users often extract screen captures of bing imagery for a variety of non-scientific purposes. It is necessary to contact the original data providers to obtain the original remote sensor data.

Esri, Inc., *ArcGIS Online Map and Geoservices*

The Esri, Inc., *ArcGIS Online Map and Geoservices* website offers information associated with basemaps, demographic maps, reference maps, specialty maps, and imagery (Figure A-7) (http://www.esri.com/software/arcgis/arcgis-online-map-and-geoservices/data-doors).

Esri Map Services

Remote sensing investigations almost always require access to other types of geospatial information. Below is a list of numerous types of image and map databases available through *ArcGIS Online* Map Services (http://www.esri.com/software/arcgis/arcgisonline/maps/maps-and-map-layers).

World Imagery Basemap: World Imagery provides ≤ 1 × 1 m satellite or aerial imagery in many parts of the world and lower resolution satellite imagery worldwide. The map includes NASA Blue Marble: Next Generation 500 × 500 m resolution imagery at small scales (above 1:1,000,000), i-cubed 15 × 15 m eSAT imagery at medium-to-large scales (down to 1:70,000) for the world, and USGS 15 × 15 m Landsat imagery for Antarctica. The map features 0.3 × 0.3 m resolution imagery in the continental United States and 0.6 × 0.6 m resolution imagery in parts of Western Europe from DigitalGlobe. In other parts of the world, 1 × 1 m resolution imagery is available from IKONOS, i-cubed Nationwide Prime, Getmapping, AeroGRID, IGN Spain,

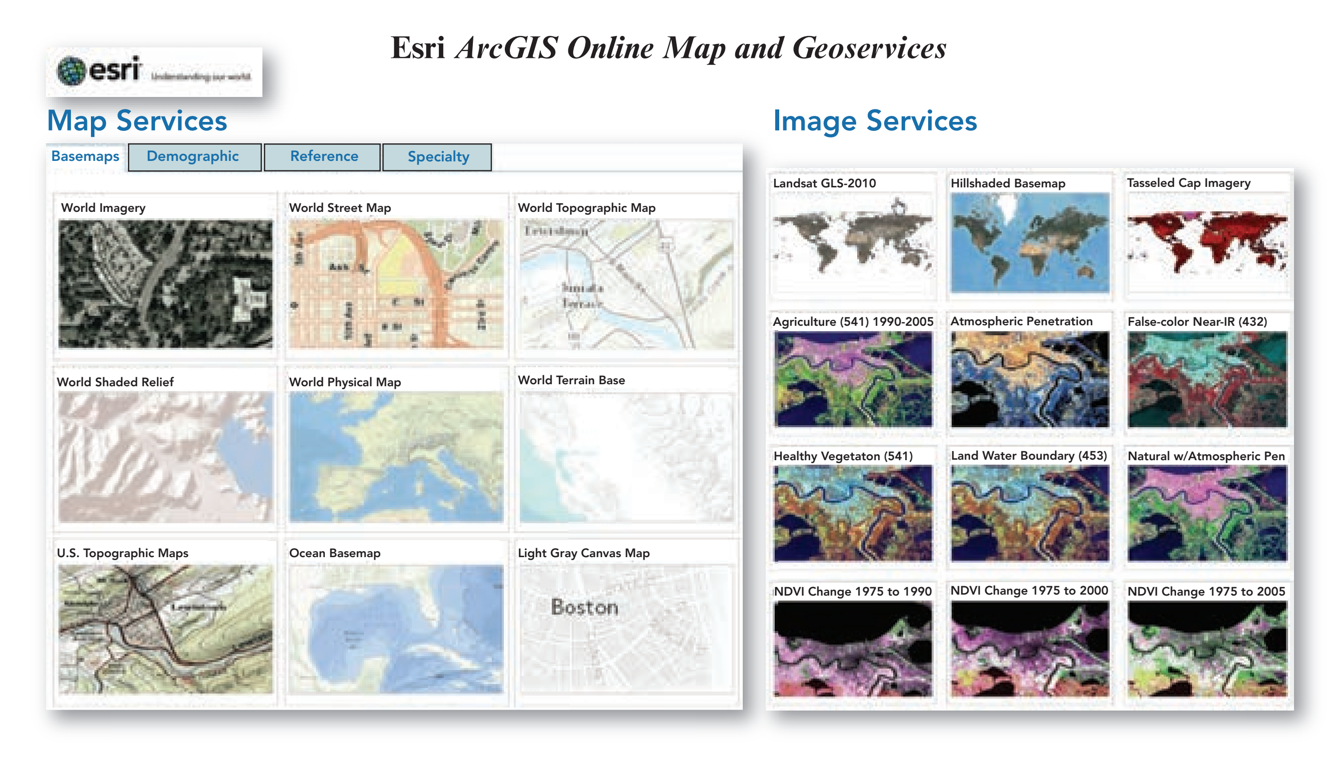


FIGURE A–7 The Esri *ArcGIS Online Map and Geoservices* interface provides access to Map Services (e.g., basemaps, demographic, reference maps, and specialty maps) and Image Services (e.g., Landsat GLS-2010, hillshaded basemap, tasseled cap imagery, and numerous Landsat band combination datasets). It also provides access to the Esri *ChangeMatters*fi program and to *MDA NaturalVue*fi satellite imagery (screenshots and images courtesy of Esri, Inc.).

and IGP Portugal. The dataset also contains data from the USDA Farm Services Agency National Agriculture Imagery Program (NAIP) imagery, USGS Digital Ortho Quarter Quad (DOQQ) imagery, and Aerials Express 0.3 to 0.6 m resolution imagery for several hundred metropolitan areas. Additionally, imagery at different resolutions has been contributed by the GIS User Community.

Other Basemaps: World Street Map, World Topographic Map, World Shaded Relief, World Physical Map, World Terrain Base, USA Topographic Maps, Ocean Basemap, Light Gray Canvas Map, National Geographic World Map, and Landsat Hill-shaded Basemap are available.

Demographic Maps: Details about the U.S. population are available, including: Average Household Size, Daytime Population, Diversity Index, Labor Force Participate Rate Median Age, Median Home Value, Median Household Income, Median Net Worth, Retail Spending Potential, Unemployment Rate, etc.

Reference Maps: World Boundaries and Places, World Boundaries and Places Alternate, World Reference Overlay, and World Transportation are available.

Specialty Maps: DeLorme World Basemap, World Navigation Charts, and Soil Survey Maps.

Esri Image Services

Esri, Inc., provides a very useful array of image services that can be of significant value to people looking for various types of remote sensor data. Below is a summary of several of the most important image services (http://www.esri.com/software/arcgis/arcgis-online-map-and-geoservices/image-services).

Landsat Global Land Survey 2010: The entire Landsat image services collection is available through *ArcGIS Online*. This service was compiled from the USGS/NASA Global Land Survey (GLS) 2010 orthorectified dataset. The dataset includes imagery from Landsat 5 TM and Landsat 7 ETM^+ at 30 × 30 m resolution, and includes all reflected energy bands, i.e., Landsat bands 1, 2, 3, 4, 5, and 7. The Landsat 7 data are corrected for Scan Line Corrector (SLC) errors. They were enhanced with radiometric correction and histogram stretching to make them more visually appealing. Landsat 8 data are now available online (http://www.esri.com/esri-news/arcnews/spring14articles/landsat-8-imagery-available-for-online-users).

Landsat Hill-shaded Basemap: Natural color, 15 × 15 m resolution, pan-sharpened Landsat imagery is orthorectified to 50 m RMSE. It is radiometric corrected and enhanced with topographic hill-shading and color balancing to produce a basemap that can be overlaid with other geospatial information using a GIS.

Landsat Color Composites 1975, 1990, 2000, 2005 and 2010: Esri provides Landsat imagery using the following band combinations (RGB = bands x, y, and z):

- False-color/Near-infrared (4,3,2),
- Agriculture (5,4,1),
- Atmospheric Penetration (7,5,4),
- Healthy Vegetation (4,5,1),
- Land/Water Boundary (4,5,3),
- Natural with Atmospheric Penetration (7,4,2),
- Shortwave Infrared (7,4,3),
- True Color/Natural Color (3,2,1), and
- Vegetation Analysis (4,5,3).

Tasselled Cap Service: This image service is based on USGS GLS 2010 orthorectified Landsat data and the tasseled cap transformation discussed in Chapter 7. Tasseled cap data are useful for monitoring vegetation development over a growing season for predicting crop maturity and yield. The Brightness image contains most of the variability in the image and is similar to panchromatic imagery. The Greenness image is useful for discriminating between plant species and quantifying plant vigor and biomass. The Wetness image is related to plant and soil moisture characteristics. The remaining bands contain noise such as haze, clouds and atmospheric affects. Brightness, Greenness and Wetness images are displayed as a color composite (RGB) where bright human-made and natural areas are red, vegetation is in shades of green, and wet or moist features are blue.

Landsat ChangeMatters®: This online service mosaics multiple-date Landsat images irrespective of the Landsat sensor that acquired them (Green, 2011). *ChangeMatters®* displays user-specified individual dates of imagery using standard band combinations, including: near-infrared color composites (e.g.,TM bands 4,3,2) useful for vegetation studies and urban analysis; natural color with atmospheric penetration (e.g., TM bands 7,4,2) best suited for analysis of vegetation and some urban studies; healthy vegetation; land/water boundary detection; and vegetation analysis providing information for agriculture and forest management. The NDVI of the two images is computed on the fly and used to document the change in NDVI between the two images (http://www.esri.com/software/landsat-imagery/viewer). Two examples of the *ChangeMatters®* user interface were presented in Chapter 12.

Esri also provides Landsat NDVI (Normalized Difference Vegetation Index) Change information associated with the following time periods: 1975 to 1990; 1975 to 2000; 1975 to 2005; 1990 to 2000; and 2000 to 2005.

MDA NaturalVue Satellite Imagery®: This dataset presents *NaturalVue®* 15 × 15 m satellite imagery of the world created by MDA Information Systems LLC. *NaturalVue®* is a natural color worldwide seamless mosaic derived over 8,600 Landsat-7 multispectral images. This single global dataset has a positional accuracy of better than 50 m RMSE.



Digital Elevation Data

Accurate digital elevation information is very important for many remote sensing and GIScience applications. A large number of commercial photogrammetric engineering firms and surveying/mapping companies provide access to high-quality digital-elevation information for a fee.

- DRG—Digital Raster Graphics (USGS at www.libremap.org).

Sometimes it is useful to analyze historical topographic information. Fortunately, the hard-copy USGS topographic maps created over many decades have been scanned to create digital raster graphics (DRGs) from the 1:24,000, 1:25,000, 1:63,360 (Alaska), 1:100,000, and 1:250,000-scale topographic map series. Coverage includes the standard USGS series quadrangle maps of the U.S. and its trusts and territories. Information inside the map neatlines is georeferenced to the surface of the Earth and fit to the Universal Transverse Mercator map projection. The horizontal positional accuracy and datum of the DRG matches the accuracy and datum of the source map. The maps were scanned at a minimum resolution of 250 dpi. Historical DRGs may be downloaded free from the U.S. Geological Survey at http://libremap.org/.

Several other important sources of digital elevation information are presented in Table A-1, including:

- GTOPO30—World digital elevation model (USGS)
- NED—National Elevation Dataset (USGS)
- Topographic and Bathymetric (USGS, NOAA)
- Topographic Change Information (USGS)
- SRTM—Shuttle RADAR Topography Mission (NASA JPL)
- ASTER Global Digital Elevation Map (METI and NASA JPL), and
- NEXTMap World 30 Digital Surface Model (Intermap, Inc.).

TABLE A-1 Digital Elevation Data.

Data	Description	Example
GTOPO30	GTOPO30 is a global digital elevation model with a horizontal grid spacing of 30 arc-seconds (approximately 1 × 1 km). It was derived from numerous sources of topographic information. It has been used extensively for regional, continental, and global studies and for numerous global change applications. Compare this GTOPO30 DEM of Mt. Kilimanjaro with the SRTM DEM to appreciate the coarse resolution of the GTOPO30 data. Source: *http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30_info*.	
NED—National Elevation Dataset	The National Elevation Dataset is the 30 × 30 m resolution elevation layer of *The National Map*. NED has evolved into a multi-resolution dataset that provides the best available elevation data, including 3 × 3 m or better LIDAR-derived elevation data. Source: *http://ned.usgs.gov*. See Gesch (2007) for a detailed description of the National Elevation Dataset characteristics.	
Topographic-Bathymetric Information	Topographic-bathymetric (topobathy) data are merged topography (land elevation) and bathymetry (water depth) data in a single product useful for inundation mapping and other applications. Topography data come from the National Elevation Dataset (NED). Bathymetry data are provided by the NOAA GEOphysical DAta System (GEODAS) .The example depicts topobathy data of a part of the Puget Sound, WA. Source: *http://topotools.cr.usgs.gov/topobathy_viewer/*.	
Topographic Change Information	The need for information on the extent of human geomorphic activity resulted in the creation of the first-ever accounting of topographic change across the U.S. The primary types of topographic changes resulting from human geomorphic activity include surface mining, road construction, urban development, dam construction, and landfills. The example shows: (left) NED DEM of a gold mine in Carlin, NV; (right) cut (blue) and fill (red) areas identified using SRTM-derived DEM. Source: *http://topochange.cr.usgs.gov/topochange_viewer/viewer.htm*.	
SRTM—Shuttle RADAR Topography Mission	SRTM data were obtained by the Space Shuttle *Endeavour* in 2000. SRTM acquired data for over 80 percent of the Earth's land surface between 60 degrees N and 56 degrees S latitude. SRTM images are available at no charge as FTP downloads. Compare the GTOPO30 DEM of Mt. Kilimanjaro with the SRTM DEM to appreciate the fine spatial resolution of the SRTM-derived DEM. Source: *http://srtm.usgs.gov/index.php*.	

TABLE A-1 Digital Elevation Data (continued).

Data	Description	Example
ASTER Global Digital Elevation Model (GDEM V2)	The Japanese Ministry of Economy, Trade, and Industry (METI) and NASA released the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 2 (GDEM V2) on October 17, 2011. ASTER GDEM coverage is from 83 degrees N to 83 degrees S latitude, encompassing 99 percent of Earth's landmass. The improved GDEM V2 adds 260,000 additional stereo-pairs, improving coverage and reducing the occurrence of artifacts. The ASTER GDEM V2 is in GeoTIFF format with 30-m postings and 1 × 1 degree tiles. ASTER GDEM V2 data are available free to users worldwide from the Land Processes Distributed Active Archive Center (LP DAAC) and J-spacesystems. The GDEM V2 coverage is shown. Source: *http://asterweb.jpl.nasa.gov/gdem.asp.*	
NEXTMap World 30 DSM (Intermap, Inc.)	Intermap Technologies, Inc., NEXTMap World 30 DSM is a combination of 90-m Shuttle Radar Topographic Mission (SRTM) v2.1 data, 30-m ASTER Global DEM v2.0, and 1-km GTOPO30 data which has been ground controlled using LiDAR data from NASA's Ice, Cloud and Land Elevation Satellite (ICESat) collection, resulting in a 25-cm RMSE dataset for vertical control of the DSM. The resulting product is a 30-m GSD DSM that covers the entire land mass of the planet.	*http://www.intermap.com/en-us/databases/world30.aspx*



Hydrography Data

Hydrologic information is critical to the successful modeling of many natural (e.g., streams) and human-made (e.g., canals) geospatial relationships. Table A-2 summarizes the availability of detailed hydrologic information from:

- NHD—National Hydrologic Database (USGS)
- EDNA—Elevation Derivatives for National Applications (USGS).



Land Use/Land Cover and Biodiversity/Habitat Data

Land use is the human use of the terrain. *Land cover* is the biophysical material present on the surface of the Earth such as vegetation, soil, water, rocks, etc. The land cover is the habitat for animals, birds, and aquatic species. Several of the more important land use/land cover and biodiversity/habitat resources are presented in Table A-3, including:

- NLCD—National Land Cover Database (USGS)
- C-CAP—Coastal Change Analysis Program (NOAA)
- GAP Analysis Program (USGS)
- NWI—National Wetlands Inventory (USFWS)

In addition, there is the *National Phenology Network*. Phenology is the study of nature's calendar. For example, the network contains information about when peach trees blossom, when alligators build their nests, when salmon swim upstream to spawn, or when corn is fully ripe. The National Coordinating Office of the Network is a resource center that facilitates and encourages widespread collection, integration, and sharing of phenology data and related information. Source: *USA National Phenology Network, 1955 East 6th Street, Tucson, AZ 85721*.

TABLE A-2 Hydrologic Information.

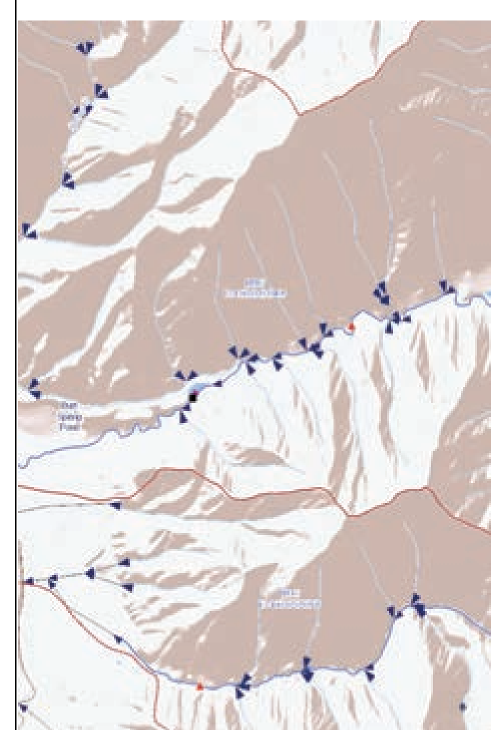
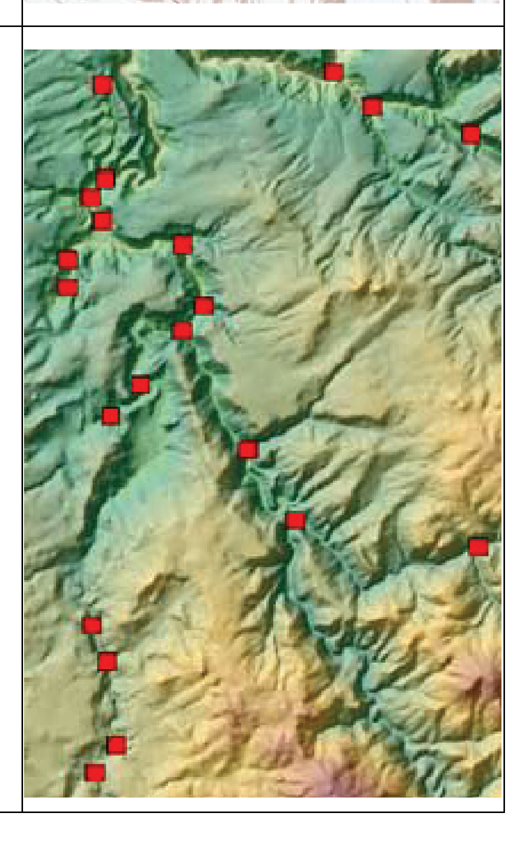
Data	Description	Example
NHD—National Hydrography Dataset	The NHD is the surface water component of *The National Map*. The NHD is a digital vector dataset easily used in GIS hydrologic analysis. It contains features such as lakes, ponds, streams, rivers, canals, dams, and stream gages. These data are designed to be used in general mapping and in the analysis of surface-water systems.The NHD contains vector-surface-water-hydrography information obtained from topographic maps and additional sources. It is available nationwide as medium resolution at 1:100,000-scale and as high resolution at 1:24,000-scale or better. In Alaska, the NHD is available at 1:63,360-scale. A few "local resolution" areas are available at varying scales. The hydrography of the U.S. is organized by drainage areas. The sub-basin 8-digit Hydrologic Unit Code (HUC) drainage area is the most practical area for high resolution NHD. Subregions (4-digit HUCs) are composed of varying numbers of subbasins. The NHD is available in an Esri personal geodatabase format known as NHDinGEO, a file-based geodatabase format, and in Esri shapefile format known as NHDGEOinShape. The NHD is organized by hydrologic units, but can be downloaded in various geographic extents. This map displays NHD data of Hobble Creek, UT. Source: *http://nhd.usgs.gov/.*	
EDNA—Elevation Derivatives for National Applications	EDNA is a multi-layered database derived from the NED, which has been hydrologically conditioned for improved hydrologic flow modeling. The seamless EDNA database provides 30 × 30 m raster and vector data layers including: Aspect Contours Filled DEM Flow Accumulation Flow Direction Reach Catchment Seedpoints Reach Catchments Shaded Relief Sinks Slope Synthetic Streamlines Hydrologically conditioned elevation data, processed to create hydrologic derivatives, can be useful in many topologically based hydrologic visualization and investigative applications. Drainage areas upstream or downstream from any location can be accurately traced, facilitating flood analysis investigations, pollution studies, and hydroelectric power generation projects. This is a map of reach catchment seedpoints. Source: *http://edna.usgs.gov/.*	

TABLE A-3 Land Use/Land Cover and Biodiversity/Habitat Information.

Data	Description	Example
NLCD—National Land Cover Database 1992, 2001, 2006, 2011	The Multi-Resolution Land Characteristics (MRLC) Consortium developed the NLCD, which is a 16 to 21 class land cover dataset of the U.S. derived from 30×30 m Landsat TM data. The NLCD state datasets are extracted from regional datasets that are mosaics of Landsat TM scenes. NLCD 1992 is a land-cover dataset. NLCD 2001 is a land-cover database composed of three elements: land cover, impervious surface, and canopy density. NLCD 2006 and 2011 are now available to the public. The land cover in Washington, DC, in 1992 is shown. Source: *http://www.mrlc.gov/.*	
C-CAP—Coastal Change Analysis Program	C-CAP is a database of land cover and land cover change for the coastal regions of the U.S. produced every 5 years. C-CAP provides inventories of coastal intertidal areas, wetlands, and adjacent uplands. The products consist of land cover maps for each date of analysis, as well as a file that highlights what changes have occurred between these dates. NOAA also produces high resolution C-CAP land cover products, for select geographies. C-CAP provides the "coastal expression" of the NLCD, and contributes to the Earth Cover layer of the National Spatial Data Infrastructure. The example shows C-CAP data of Pearl Harbor, HI. Source: *http://coast.noaa.gov/digitalcoast/data/ccapregional/?redirect=301ocm.*	
GAP Analysis Program	The GAP Analysis Program is part of the National Biological Information Infrastructure (NBII). The goal is to keep common species common by protecting them *before* they become threatened. GAP activities focus on the creation of state and regional databases and maps that depict patterns of land management, land cover, and biodiversity. These data can be used to identify "gaps" in conservation where an animal or plant community is not adequately represented on the existing network of conservation lands. The example depicts the U.S. GAP land cover map which contains 551 ecological systems and modified ecological systems. Source: *http://gapanalysis.usgs.gov/.*	
NWI—National Wetlands Inventory	The U.S. Fish & Wildlife Service's National Wetlands Inventory provides current geospatial data on the status, extent, characteristics, and functions of wetland, riparian, deepwater, and related aquatic habitats in priority areas to promote the understanding and conservation of these resources. The NWI serves on-line map information for 82% of the conterminous United States, 31% of Alaska, and 100% of Hawaii. The goal is to complete and maintain a seamless digital wetlands dataset for the entire nation that will become the Wetlands Data Layer of the NSDI. The NWI example is of Crayvik, FL. Source: *http://www.fws.gov/wetlands/.*	



Road Network and Population Demographic Data

Topologically correct road-network information is indispensable for navigation, business decision making and for use in various remote sensing investigations. The U.S. Census Bureau provides the MAF/TIGER—Master Address File/Topologically Integrated Geographic Encoding and Reference System free of charge to the public. Details are provided in Table A-4.

The Census Bureau also provides population demographic information aggregated to various spatial enumeration districts (Table A-4).

LandScan, created by the Oak Ridge National Laboratory, predicts the spatial distribution of people at different times of day anywhere in the world at a spatial resolution of 1 × 1 km (Table A-4).



Remote Sensor Data—Public

Remote sensing data are available from orbital and aerial platforms and sensors. Table A-5 provides information about some of the public remote sensing systems commonly used to collect data.

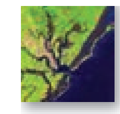
- ALI—Advanced Land Imager (NASA)
- ASTER—Advanced Spaceborne Thermal Emission and Reflection Radiometer (NASA)
- AVHRR—Advanced Very High Resolution Radiometer (NOAA)
- AVIRIS—Advanced Visible Infrared Imaging Spectrometer (NASA)
- Declassified Satellite Imagery (USGS)
- DOQ—Digital Orthophoto Quadrangles (USGS)
- Hyperion—Hyperspectral Imager (NASA)
- Landsat Multispectral Scanning System (MSS), Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM^+), and Landsat 8 (NASA/USGS)
- LiDAR—Light Detection and Ranging (USGS)
- MODIS—Moderate Resolution Imaging Spectrometer (NASA)
- NAIP—National Agriculture Imagery Program (USDA)
- Suomi—NPOESS Preparatory Project (NPP) (NOAA).



Remote Sensor Data—Commercial and International

Remote sensor data available from selected commercial and international vendors are provided in Table A-6.

- CASI-1500, SASI-600, MASI-600 and TASI-600 (ITRES, Inc.)
- EROS A and EROS B (ImageSat Intl., Inc.)
- GeoEye-1 and GeoEye-2 (DigitalGlobe, Inc.)
- HyMap Hyperspectral imagery (HyVista, Inc.)
- IKONOS-2 (DigitalGlobe, Inc.)
- Indian IRS A-D, CartoSat 1-3, ResourceSat 1-2 (www.isro.gov.in)
- Korean KOMPSAT 1-5 (http://www.kari.re.kr)
- Pictometry (EagleView Technologies, Inc.)
- Pleiades-1 and Pleiades-2 (www.astrium-geo.com)
- QuickBird (DigitalGlobe, Inc.)
- RapidEye (RapidEye, Inc.)
- Sentinel-2 (European Space Agency)
- SPOT 1-6 (www.astrium-geo.com)
- WorldView-1, WorldView-2, WorldView-3 (Digital Globe, Inc.).



References

Carswell, W. J., 2011, *National Geospatial Program*, Washington: U.S. Geological Survey, 2 p.

Gesch, D. B., 2007, Chapter 4: “The National Elevation Dataset,” in Maune, D. F., *Digital Elevation Model Technologies and Applications—The DEM Users Manual* (2nd ed.), Bethesda: American Society for Photogrammetry & Remote Sensing, 99–118.

Green, K., 2011, “Change Matters,” *Photogrammetric Engineering & Remote Sensing*, 77(4):305–309.

Moe, D. and R. Longhenry, 2013, “Metrically Preserving the USGS Aerial Film Archive,” *Photogrammetry Engineering & Remote Sensing*, 79(3):225–228.

USGS NGP, 2014, *National Geospatial Program*, http://www.usgs.gov/ngpo.

TABLE A-4 Road Network and Population Demographic Information.

Data	Description	Example
MAF/TIGER/ Line	TIGER/Linefi Shapefiles are spatial data extracts from the U.S. Census Bureau MAF/TIGER database (Master Address File/Topologically Integrated Geographic Encoding and Referencing System), containing features such as topologically correct roads, railroads, and rivers, as well as legal and statistical geographic areas. They are available to the public at no charge and are typically used to provide the digital base map for GIS software.The TIGER/Linefi Shapefiles do not include demographic data, but they contain geographic entity codes that can be linked to the Bureau's demographic data, available on American FactFinder described below. The TIGER /Line Shapefiles are provided in four types of coverages: County-based State-based Nation-based American Indian Area-based Source: *http://www.census.gov/geo/maps-data/data/tiger.html.*	
2010 Census Population Demographics	The 2010 Census reported that 308.7 million people live in the U.S., a 9.7 percent increase from the 2000 Census population of 281.4 million. Below are some of the most important 2010 United States Census products (tabular and feature-based). *American FactFinder (AFF)* The AFF is a data access system that will find and retrieve geographic information and plot the information on an interactive map. Source: *http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml.* *Cartographic Boundary Files* Generalized, digital files suitable for use with GIS in small-scale thematic mapping. Source: *http://www.census.gov/geo/maps-data/data/tiger-cart-boundary.html.* *Census 2010 County Block Maps* Color maps, showing census blocks, voting districts, and other feature details. Source: *http://www.census.gov/geo/maps-data/maps/block/2010/.* *Voting District/State Legislative District Reference Maps* Color maps, showing voting districts and/or state legislative districts. Source: *http://www.census.gov/geo/maps-data/maps/reference-sld.html.*	
LandScan Population Distribution Modeling	*LandScan* was developed by Oak Ridge National Laboratory (ORNL) through funding by the DOD and the DOE. At approximately 1 × 1 km, *LandScan* is the highest resolution global population distribution data available and represents an ambient population (average over 24 hours). The *LandScan* algorithm uses spatial data and image analysis technologies and a multivariable dasymetric modeling approach to disaggregate census counts within an administrative boundary. *LandScan* population distribution models are tailored to match the data conditions and geographical nature of individual countries and regions. The *LandScan* dataset is available free to U.S. Government agencies. Others, such as educational institutions or research scientists, must register with *LandScan* to determine license fees (if any) for obtaining *LandScan* resources. The data are distributed in both an ESRI grid format and an ESRI binary raster format. The dataset has 20,880 rows and 43,200 columns covering north 84 degrees to south 90 degrees and west 180 degrees to east 180 degrees. Source: *http://www.ornl.gov/sci/landscan/index.shtml.*	

TABLE A-5 Remote Sensor Data—Public.

Data	Description	Example
ASTER—Advanced Spaceborne Thermal Emission and Reflection Radiometer	ASTER is carried onboard NASA's *Terra* satellite. ASTER began data collection in 2000 and consists of three sensors that obtain images in multiple spatial resolutions (15 × 15, 30 × 30, and 90 × 90 m). ASTER is the only high-resolution imaging sensor on *Terra*. The primary goal of the ASTER mission is to obtain high-resolution image data in 14 channels (including thermal infrared) over the entire land surface, as well as black and white stereo images. With a revisit time of between 4 and 16 days, ASTER provides repeat coverage of changing areas on Earth's surface. This is an ASTER image of the Patagonia Glacier in Chile obtained on May 2, 2000. Source: *http://asterweb.jpl.nasa.gov/gdem.asp*.	
AVHRR—Advanced Very High Resolution Radiometer	The Advanced Very High Resolution Radiometer (AVHRR) is an optical multispectral scanner flown aboard NOAA orbiting satellites. The instrument measures reflected sunlight and emitted radiation (heat) from Earth in the visible (Channel 1), near-infrared (Channel 2), and thermal infrared (Channels 3, 4, and 5) regions of the electromagnetic spectrum. There is fairly continuous global coverage since June 1979, with morning and afternoon acquisitions available. The resolution is 1.1 × 1.1 km at nadir. This is an image of the Gulf Stream off of North and South Carolina. Source: *http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/AVHRR*.	
AVIRIS—Airborne/Visible Imaging Spectrometer	AVIRIS is a unique optical sensor that delivers calibrated images of the upwelling spectral radiance in 224 contiguous spectral channels (bands) with wavelengths from 400 to 2,500 nanometers. The main objective of the AVIRIS remote sensing system is to identify, measure, and monitor constituents of the Earth's surface and atmosphere based on molecular absorption and particle scattering signatures. Research with AVIRIS data is predominantly focused on understanding processes related to the global environment and climate change. This is an AVIRIS image of the lower portion of the San Francisco Bay, California. Source: *http://aviris.jpl.nasa.gov/*.	
Declassified Satellite Imagery	Almost 90,000 declassified satellite images are maintained in the USGS EROS declassified satellite image archive. The images were captured by a variety of intelligence satellites, including CORONA, ARGON, LANYARD, KH-7, and KH-9 between 1960 and 1980. Coverage is global, but the geographic distribution is uneven. Some images are high resolution (although not georeferenced). This is an image of the western edge of Dakar, Senegal, Africa, acquired by the KH-7 satellite in 1966. Source: *http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/Declassified_Satellite_Imagery_-_2*.	

TABLE A-5 Remote Sensor Data—Public (continued).

Data	Description	Example
DOQ—Digital Orthophoto Quadrangles	A Digital Orthophoto Quadrangle (DOQ) is a computer-generated image of an aerial photograph in which the image displacement caused by terrain relief and camera tilt has been removed. The DOQ combines the image characteristics of the original photograph with the georeferenced qualities of a map. DOQs can be black-and-white (B/W), natural color, or color-infrared (CIR) images with 1 × 1 m ground resolution. They cover an area measuring 3.75-minutes longitude by 3.75-minutes latitude or 7.5-minutes longitude by 7.5-minutes latitude. A CIR DOQ of downtown Washington, DC, is shown. Source: *https://lta.cr.usgs.gov/DOQs*.	
Landsat—MSS, TM ETM^{+}, Landsat 8	The first two Earth Resources Technology Satellite (ERTS) were launched in 1972 and 1975. They were subsequently renamed Landsat satellites. Additional Landsat satellites were launched in 1978, 1982, 1984, 1993 (did not achieve orbit), 1999 and 2013. The Landsat satellites have carried a variety of sensors, including the Multispectral Scanning System (MSS; 80 × 80 m), Thematic Mapper (TM; 30 and 60 m), and Enhanced Thematic Mapper (ETM^{+}; 15, 30 and 60 m). Landsat 8 was launched on February 11, 2013, with its Operational Land Imager; 15 and 30 m) and Thermal Infrared Sensors (100 m). All Landsat data acquired from 1972 are available at no charge from the USGS. A Landsat 7 image draped over an SRTM digital elevation model of southern Malawi, Africa, near the Mozambique border is shown. Source: *http://landsat.gsfc.nasa.gov/images/archive/f0005.html*.	
LiDAR— Light Detection and Ranging	LiDAR data collection is one of the most important sources of digital surface models that provide detailed information about the *x*,*y*, and *z*-location of masspoints for all features on the surface of the Earth. With accurate processing, the masspoints can be edited to produce bare-earth digital terrain models. Additional processing can yield biophysical information about the vegetation. LiDAR can also be used to obtain bathymetric information in shallow clear water. *Click* is the USGS Center for LiDAR Information Coordination and Knowledge at *http://lidar.cr.usgs.gov/*.	
MODIS—Moderate Resolution Imaging Spectrometer	MODIS is flown on both the NASA *Aqua* and *Terra* satellites. It collects data in 36 spectral bands at three spatial resolutions: 250 × 250, 500 × 500, and 1,000 × 1,000 m. MODIS collects imagery of the entire Earth every 1 to 2 days. The data are excellent for regional, continental, and global analyses of relatively small-scale problems such as dynamic land, ocean, ice, and atmospheric processes and global environmental change. Source: *http://modis.gsfc.nasa.gov/gallery/#*.	

TABLE A-5 Remote Sensor Data—Public (continued).

Data	Description	Example
NAIP—National Agriculture Imagery Program	The NAIP acquires imagery during the agricultural peak growing seasons for the conterminous United States. The "leaf-on" orthophoto images are at 1 × 1 m or 2 × 2 m spatial resolutions. The 1 × 1 m imagery provides updated digital orthophoto imagery. The 2 × 2 m imagery supports the USDA Farm Services programs that require current imagery acquired during the agricultural growing season but do not require a higher resolution. Included in this imagery category is 1 × 1 m data acquired through certain state programs as "leaf-off." Source: *http://www.fsa.usda.gov/FSA/apfoapp?area=home&subject=prog&topic=nai*.	
Suomi—NPOESS Preparatory Project	The Suomi NPOESS Preparatory Project (NPP) satellite was launched on October 28, 2011. It has five imaging systems, including: Visible Infrared Radiometer Suite (VIIRS) consisting of 22 channels between 0.4 and 12 μm; Advanced Technology Microwave Sounder (ATMS); Cross-track Infrared Sounder (CrIS); Ozone Mapping and Profiler suite (OMPS); and Clouds and the Earth's radiant Energy System (CRES) radiometer. This image of the continental U.S. at night is a composite assembled from data acquired by the Suomi NPP satellite in April and October 2012. The image was made possible by the "day-night band" of the VIIRS, which detects light in a range of wavelengths from green to near-infrared and uses filtering techniques to observe dim signals such as city lights, gas flares, auroras, wildfires and reflected moonlight. Credit: NASA Earth Observatory/NOAA NGDC. Source: *http://npp.gsfc.nasa.gov/*.	

TABLE A-6 Remote Sensor Data—Commercial and International.

Data	Description	Source
CASI-1500 SASI-600 MASI-600 TASI-600	ITRES, Inc. of Canada markets several airborne optical hyperspectral remote sensing systems: [CASI-1500 (0.37 – 1.05 μm) and SASI-600 (0.95 – 2.45 μm)], and hyperspectral thermal infrared systems: [MASI-600 (4 – 5 μm) and TASI-600 (8 – 11.5 μm)]. Typical ground spatial resolution is between 1 and 10 m.	*www.itres.com*
EROS A EROS B	EROS A was launched on December 5, 2000, with a single 1.9 × 1.9 m panchromatic band (0.5 – 0.9 μm). EROS B was launched on April 25, 2006, with the identical sensor.	*www.imagesatintl.com*

TABLE A-6 Remote Sensor Data—Commercial and International (continued).

Data	Description	Source
GeoEye-1 **GeoEye-2**	GeoEye-1 was launched on September 6, 2008, with a 0.41 × 0.41 m panchromatic band and four 1.65 × 1.65 m multispectral bands. GeoEye-2 has a 0.25 × 0.25 m panchromatic band and four 1.65 × 1.65 m multispectral bands. GeoEye, Inc. merged with DigitalGlobe, Inc. in February, 2013.	*www.digitalglobe.com*
HyMap	Airborne HyMap hyperspectral data are recorded in 128 bands in the region from 450 – 2,480 nm. Typical ground resolution is between 2 and 10 m.	*www.hyvista.com*
IKONOS-2	IKONOS-2 was launched on April 27, 1999. It has a 0.82 × 0.82 m panchromatic band and four 3.2 × 3.2 m multispectral bands. GeoEye, Inc. merged with DigitalGlobe, Inc. in February, 2013.	*www.digitalglobe.com*
Indian **IRS-1A** **IRS-1B** **IRS-1C** **IRS-1D** **CartoSat-1** **CartoSat-2** **CartoSat-2A** **Cartosat-2B** **Cartosat-3** **ResourceSat-1** **ResourceSat-2**	The Indian Space Research Organization (ISRO) launched IRS-1A on March 17,1988; IRS-1B on August 29, 1991; IRS-1C on December 28, 1995; and IRS-1D on September 29, 1997. CartoSat-1 was launched on May 5, 2005, with a 2.5 × 2.5 m panchromatic band on two cameras. CartoSat-1A will have 1.25 × 1.25 m pan band, 2.5 × 2.5 m multispectral bands, and 30 × 30 m hyperspectral bands. CartoSat-2 was launched on January 10, 2007, with a 1 × 1 m pan band. CartoSat-2A was launched on April 28, 2008, with the same specifications. CartoSat-2B was launched on July 12, 2010, with a 1 × 1 m pan band. CartoSat-3 will have a 0.25 × 0.25 pan band. ResourceSat-1 was launched on October 17, 2003, with three sensors recording imagery at 5.8 × 5.8 m, 23.5 × 23.5 m, and 56 × 56 m spatial resolutions. ResourceSat-2 was launched on April 20, 2011, with swath width improvements.	*www.isro.gov.in*
Korean **KOMPSAT-1** **KOMPSAT-2** **KOMPSAT-3** **KOMPSAT-4** **KOMPSAT-5**	The Korean Aerospace Research Institute (KARI) launched KOMPSAT-1 on December 21, 1999, with 6 × 6 m panchromatic band. It was retired in 2008. KOMPSAT-2 was launched on July 27, 2006, and has a 1 × 1 m panchromatic band and four 4 × 4 m multispectral bands. KOMPSAT-3 was launched on May 18, 2012, with a 0.8 × 0.8 m panchromatic band and a 4 × 4 m VNIR multispectral sensor. KOMPSAT-3A will have sensors similar to KOMPSAT-3 except for a 5.5 × 5.5 m thermal infrared band in the region 3 – 6 μm. KOMPSAT-5 is to have a Synthetic Aperture Radar (SAR) that can acquire data at 1 × 1 m spatial resolution.	*http://www.kari.re.kr (Korean)*
PICTOMETRY	Pictometry International and EagleView Technologies merged in 2013. Pictometry imagery consists of a single Nadir (vertical) image and four low-oblique images (North, East, South, and West) collected at each exposure station. Natural color and color-infrared imagery can be collected. The flightlines overlap by 20% to 30% resulting in a great diversity of views of each terrain object.	*www.eagleview.com/Products/ImageSolutionsAnalytics/PictometryImagery.aspx*

TABLE A-6 Remote Sensor Data—Commercial and International (continued).

Data	Description	Source
Pleiades-1 **Pleiades-2**	Pleiades-1 was launched on December 17, 2011, and Pleiades-2 was launched on December 2, 2012. Each satellite sensor has a 0.5 × 0.5 m panchromatic band and four multispectral bands at 2 × 2 m spatial resolution.	*www.astrium-geo.com/pleiades/*
QuickBird	DigitalGlobe, Inc. launched QuickBird on October 18, 2001, with a 0.65 × 0.65 m panchromatic band and four 2.62 × 2.62 m multispectral bands.	*www.digitalglobe.com*
RapidEye	Five identical RapidEye satellites were launched on August 29, 2008. Each RapidEye sensor system has five multispectral bands at 6.5 × 6.5 m spatial resolution (5 × 5 m when processed to become orthoimagery).	*www.rapideye.com*
Sentinel-2	The European Space Agency (ESA) is developing a new family of missions called Sentinels specifically for the operational needs of the Copernicus programme. Sentinel-2 will have a Multi-Spectral Instrument (MSI) with a total of 13 bands: 4 VNIR bands (400 – 750 nm) at 10 × 10 m, 6 SWIR bands (1,300 – 3,000 nm) at 20 × 20 mm and 3 bands at 60 m spatial resolution dedicated to atmospheric correction and cloud screening.	*www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2*
SPOT 1 **SPOT 2** **SPOT 3** **SPOT 4** **SPOT 5** **SPOT 6** **SPOT 7**	SPOT 1, 2, and 3 are retired and SPOT 4, 5, and 6 are operational. Archived SPOT 1, 2, and 3 data have a 10 × 10 m panchromatic band and three 20 × 20 m multispectral bands. SPOT 4 has a 10 × 10 m panchromatic band, three 20 × 20 m multispectral bands, and one 20 × 20 m SWIR band. SPOT 5 has a 2.5 × 2.5 m panchromatic band, three 10 × 10 m multispectral bands and one 20 × 20 m SWIR band. SPOT 4 and SPOT 5 have a unique 1.15 × 1.15 km Vegetation sensor with four multispectral bands. SPOT 6 was launched on September 9, 2012, with a 1.5 × 1.5 m panchromatic band and four multispectral bands at 6 × 6 m spatial resolution. SPOT 7 was launched on June 30, 2014, with the same payload as SPOT 6.	*www.astrium-geo.com/en/147-spot-6-7-satellite-imagery*
WorldView-1 **WorldView-2** **WorldView-3**	WorldView-1 was launched on September 18, 2007, with a 0.5 × 0.5 m panchromatic band. WorldView-2 was launched October 8, 2009, with a 0.46 × 0.46 m panchromatic band and eight multispectral bands at 1.85 × 1.85 m spatial resolution. WorldView-3 was launched on August 13, 2014, with a 0.31 × 0.31 m panchromatic band, eight 1.24 × 1.24 m multispectral bands, and eight 3.7 × 3.7 m SWIR bands. In June, 2014, DigitalGlobe received permission from the U.S. Department of Commerce to sell WorldView-3 imagery at their best available resolution: 0.25 × 0.25 m panchromatic and 1 × 1 m multispectral.	*www.digitalglobe.com*

Index

E

G

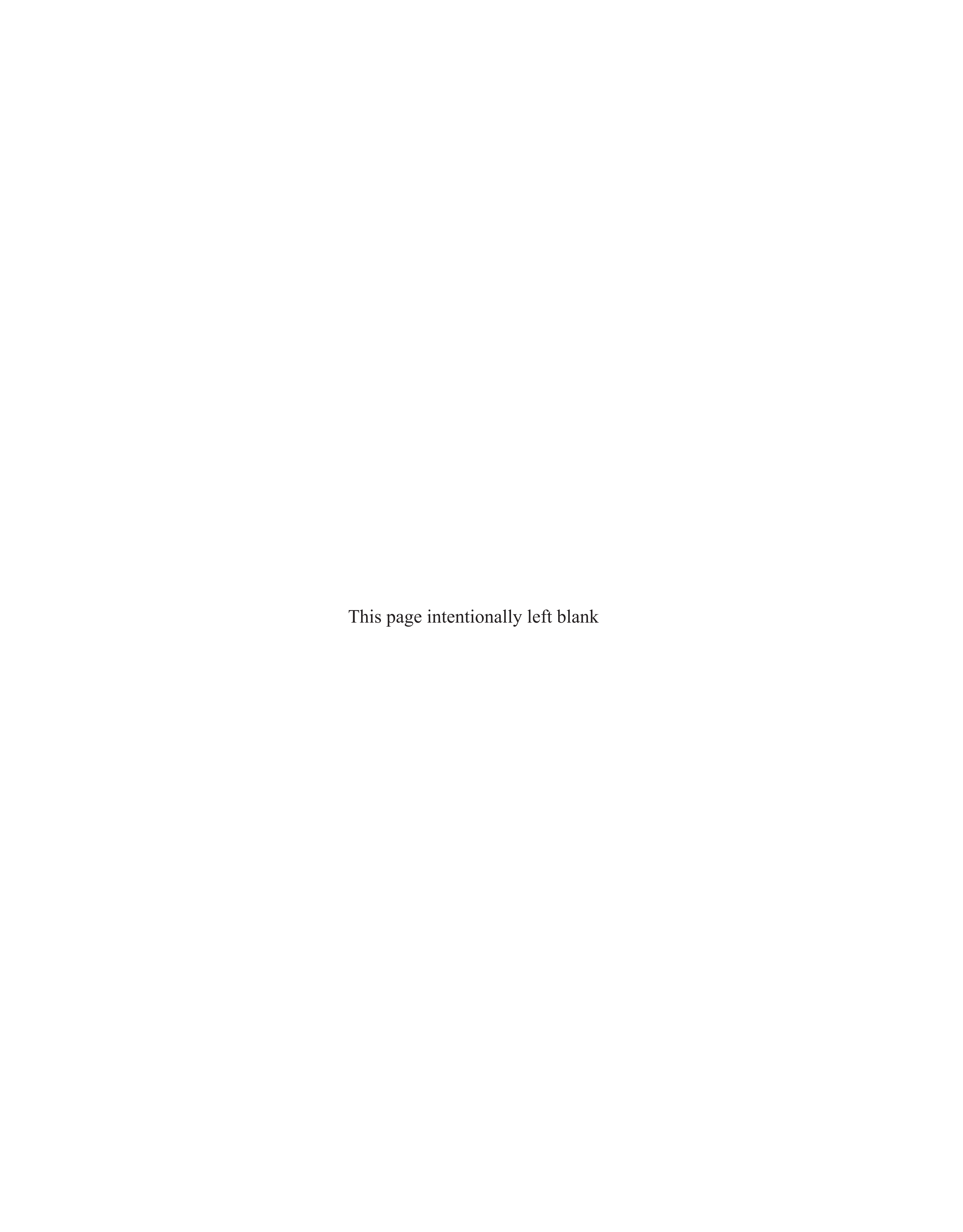
H

M

S

T

U



This page intentionally left blank

INTRODUCTORY
DIGITAL IMAGE PROCESSING
A Remote Sensing Perspective

4th Edition

John R. Jensen

Digital image processing of remotely sensed data is a rapidly developing technology used by scientists and agencies throughout the world to obtain information for managing natural resources and many other applications. *Introductory Digital Image Processing: A Remote Sensing Perspective, 4th Edition* introduces how remote sensing imagery is collected. It then presents the algorithms used to radiometrically and geometrically prepare digital imagery for further analysis. Great emphasis is then placed on the use of state-of-the-art digital image processing techniques to enhance single-date images and to extract useful thematic information from the imagery (e.g., land cover). The book then reviews algorithms for digital change detection using multiple dates of imagery. Change information is used to identify manmade and physical processes at work to solve real-world problems. The final chapter describes how to assess the accuracy of thematic map products derived from remote sensing.

About the Cover: The top image is an artist's rendition of the Landsat 8 Earth observation satellite in orbit over the southeastern United States. Since 1972, the Landsat satellites have provided a continuous record of Earth's changing land surfaces. It is the world's largest and most sustained collection of Earth observation imagery. The lower left image is a Landsat 8 scene of the Cambridge Gulf in Northwestern Australia. *Geoscience Australia* enhanced the image to display sediment and nutrient patterns in the tropical estuary area, and the complex patterns and conditions in the vegetated areas. The lower right image is a natural color Landsat 8 image of the salt evaporation ponds in the southern part of San Francisco Bay, California. The image was processed to highlight the ponds, which are designed to produce salts from sea water. Salt ponds range in color from blue green to deep magenta, where the color indicates the salinity of the ponds and the type of microorganisms breeding in it. *Introductory Digital Image Processing* explores the tools and techniques used to create and analyze images such as these.

